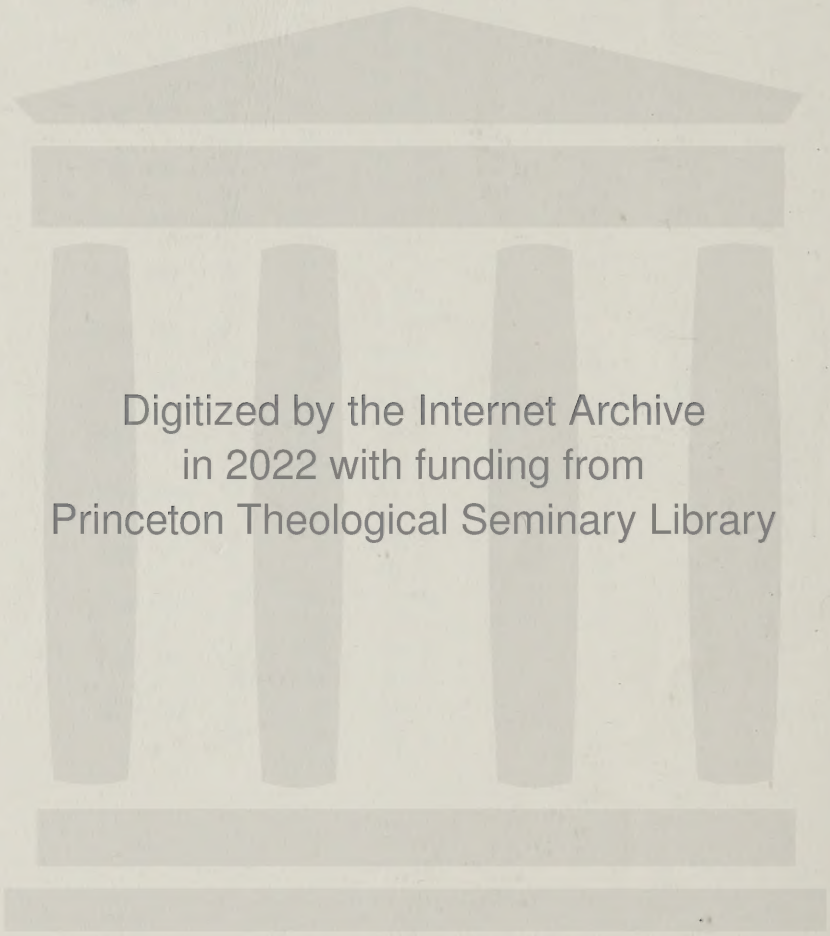


VOL. XXXIII
NO. 6

PSYCHOLOGICAL REVIEW PUBLICATIONS

WHOLE NO. 153
1924

Psychological Monographs

A Group Intelligence Scale for Primary Grades

BY
FORREST ALVA KINGSBURY

PSYCHOLOGICAL REVIEW COMPANY
PRINCETON, N. J.

AGENTS: G. E. STECHERT & CO., LONDON (2 Star Yard, Carey St., W.C.)
PARIS (16, rue de Condé)

PREFATORY NOTE

The interval of three years which has elapsed since the completion of this study and the publication of the scale would suggest certain possible revisions of phrasing; but since no definite reason has appeared for altering materially the statements made herein, the writer has preferred to let it appear substantially in its original form. The revised grade-norms cited on page 33 based on returns from various quarters since the scale was taken over for publication by the Bureau of Educational Research (and since distributed by the Public School Publishing Company) represent the principal addition.

The writer welcomes this opportunity to express his real indebtedness to numerous teachers and associates in the University of Chicago for very helpful criticisms and suggestions, and particularly to Professor Frank N. Freeman of the School of Education, under whose direction the study was made, and to Professor Harvey Carr, of the Department of Psychology. He is also grateful to the many school administrators and teachers whose courteous cooperation has made this investigation possible.

A GROUP INTELLIGENCE SCALE FOR PRIMARY GRADES

Introduction

Due both to the stimulus of the Army experience and to the demand for economy, the tendency of the mental testing movement in the schools has come to be definitely in the direction of the development and refinement of group tests. Various workers have designed more or less satisfactory tests for the upper grades and high schools, but until comparatively recently the primary grades have been neglected. This has doubtless been due to the difficulty of devising tests for pupils who have little or no skill in reading. Nevertheless, if there is any stage in the school course where there is need for tests of that general mental ability to which the term "general intelligence" is customarily applied, it certainly is in the first three grades of the elementary school, where the dominating school habits and attitudes are acquired. The need of lessening the amount of non-promotion with its attendant expensiveness and discouragement, the desirability of putting each child in homogeneous groups where he can make progress with maximal efficiency, the economy of selecting early the exceptionally gifted and exceptionally slow children and giving them appropriate treatment at the beginning of their school course, these and many other almost universally recognized needs call for some sort of instrument for selecting different grades of ability which will be suitable for the years before the child has gained sufficient skill in reading and writing to use tests which presuppose those abilities. The Binet test, with its various revisions, has served as such an instrument. This is not the place to discuss the precise degree of efficiency with which it has served its purpose. But no one will deny that the expenditure of time, and hence of money, called for in administering the test to not more than ten or twelve pupils per day per examiner is prohibitive of its use in most places, unless it be with a limited number of cases, usually exceptional children, whose superiority or inferiority is so conspicuous that a mental test is, to a certain extent, superfluous. While most teachers have some acquaintance with the general purpose of the mental testing movement, and perhaps some vague notion of its nature, the large

majority of teachers (keeping in mind particularly the vast number in rural and village schools) probably have never seen such a test given, and have neither the equipment, training, nor time to give it. If satisfactory tests can be devised, their most conspicuous advantage will be the economy of time and cost, and their almost certain influence upon the introduction of mental testing, and eventually of other scientific educational practices, into a large number of schools which have hitherto been little affected by scientific movements.

The fact that a number of workers have, within recent years, undertaken to devise group test scales for the primary grades, has not deterred the writer from undertaking a similar task. There is, it would seem, a need for several good group tests for these grades, for two reasons: first, in order that as wide a variety as possible of types of material may be compared, to determine their relative value for this purpose, and second, because, in the writer's firm belief, no one test should be relied upon to serve as a sole means for determining the disposition of a child's case; rather, it would seem, will the desirable practice be to administer, during the first year or two of the school career, a number of tests of proved value, in order to counteract the variability of conditions under which each test is given which may prevent the child from giving an adequate account of himself, and in order to disclose his peculiarities of response to different sorts of problems.

Among the requirements of a good group intelligence scale for primary grades are the following:

1. It must permit of being given to a group of pupils at once, preferably to an entire room.

2. It should, if possible, dispense entirely with reading and writing content, since varying school attainment in these subjects is likely to obscure differences in native endowment.

3. It should be brief, to avoid fatigue and waning of interest.

4. It should be easy and uniform, both to administer and to score, to permit its use by relatively untrained examiners.

5. Instructions should be simple enough for all children to grasp, without such elaboration or repetition as will cause loss of interest among the quicker pupils.

6. It should be interesting to the pupils, utilizing the play motive, in order to claim and hold attention at a maximum.

7. It should show as high a correlation as possible with all other available criteria of general intelligence.

8. Norms for age, grade, sex, etc., should be based on a sufficiently large number of cases to be representative and reliable.

Of these points in relation to the scale herein described, more will be said later.

Procedure

The procedure followed in formulating and standardizing this scale may be summarized in the following steps:

1. Devising the tests.
2. First try-out to eliminate useless material and formulate instructions.
3. Arranging material for preliminary group test.
4. Formulating instructions, determining time, etc. (technique).
5. Giving preliminary group test.
6. Scoring papers and tabulating results to facilitate computations.
7. Computation of correlations for
   a—Individual tests
   b—Parts within tests
   c—Total score
   d—Various combinations of parts with such criteria as
      a—Binet mental age
      b—Teachers' estimates of intelligence.
8. Determining final make-up of scale.
9. Preparing scale for giving final form in large numbers.
10. Giving test to a considerable number of primary grade groups.
11. Scoring papers and tabulating results.
12. Computation of norms (median, quartile, decile) for each
   a—Grade
   b—Age
   c—Sex.
13. Obtaining from teachers analytical comments on cases where test score and teachers' estimates disagree.

14. Analysis of scores, norms, teachers' comments, and peculiarities of response.

These steps will be considered in the order named.

The Tests

In devising the tests to be tried out, three principles were consciously followed: (1) to use no tests others are using; (2) to adapt for non-verbal form such tests as have been shown in the past to have high value as criteria of general intelligence, such as the Opposites test, Series Completion test, Analogies test, etc.; (3) to use tests calling for judgment and thought activities rather than tests of perception, simple association, memory, etc.

The tests first devised were ten in number. Those which were retained and used in group form are illustrated in Plates I to VI and the instructions accompanying them are given in Appendix II.

a. Simple Directions test (ten parts). Some of these parts were later combined with the Right Answers test.

b. Opposites test (twenty parts).

c. Associated Objects test (twenty parts).

d. Right Answers test (a miscellaneous group of fifteen single tests, most of which are described later).

e. Series Completion test (twenty parts).

f. Analogies test (twenty parts).

g. Form test, or "Dissected Blocks" test, based on the square (ten parts).

h. Form test, based on the circle (ten parts).

i. Path tracing test, in which ten interweaving lines, distinguished by a simple symbolic mark (circle, square, cross, etc.) at the beginning of each line, were to be traced across the page, and a corresponding symbol marked at the end of each.

j. Domino test (sixteen parts), a variety of "Series Completion" test, in which a row of five dominoes were shown, one blank, to be filled in with dots in such a way as to complete the series. Somewhat similar tests are included in the Series Completion test.

These tests were tried out on individual children, about twenty in number, taken at random from the first three grades, for the

purpose of observing typical reactions to the various tests, eliminating valueless material, discovering and correcting specific defects, formulating instructions, and ascertaining the relative difficulty of the various parts within each test. The children were taken separately, usually in two half-hour periods; and while the test was made as informal as possible, a record of the responses and time was kept. In most cases the child was asked to respond by pointing to the correct figure or drawing. Where the response required drawing, a ground glass or strip of paper was laid over the test, upon which the child marked. Thus only one copy of the preliminary form was used. The directions as originally prepared were gradually amended until a clear, simple, and unambiguous form of expression was reached.

From this series of tests a tentative group was prepared, by selection and combination, for use as a preliminary group test. Parts that were ambiguous, too easy, or apparently too difficult for any child to solve, were omitted. The Path-tracing test (i) was eliminated, because, both from analysis of the test and from observation of children's reactions to it, it seemed to be primarily a sensory-motor test, measuring motor control rather than intelligence. The Domino test (j) was eliminated, as duplicating the other Series Completion test (e). Some of the Simple Directions tests were combined with the Right Answers tests. The Form test based on the circle was eliminated in favor of the one based on the square, but several parts of the former were carried over and adapted to the square form.

The Preliminary Group Test

This left six tests which are illustrated in Plates I to VI inclusive. Plates I to IV, it may be remarked, represent the scale in its final form, while Plates V and VI show in reduced form the materials which, together with the other four tests, were used in the preliminary group test. This material was duplicated by mimeograph and stapled in booklets of six sheets (three tests) each, so the preliminary group test was given in two parts. The first part included:

a. Associated Objects test (fourteen parts, exclusive of the

PLATE I
(Reduced about two-fifths)

For Grades 1, 2, 3 and 4

UNIVERSITY OF ILLINOIS
Urbana, Illinois
BUREAU OF EDUCATIONAL RESEARCH

Kingsbury Primary Group Intelligence Scale, Form A

Devised by Forrest A. Kingsbury

Name ..... Boy or Girl ..... Age last birthday ..... Next birthday will be ..... 192..... Grade .....

Date ..... City ..... State .....

School ..... Teacher .....

Write Pupil's Scores Here	
1-Rights x 3	Mental Age .....
2-(R-½W) x 2	
3-Rights x 2	Chronological Age .....
4-Rights x 2	
Total Score	Intelligence Quotient .....

1. RIGHT ANSWERS

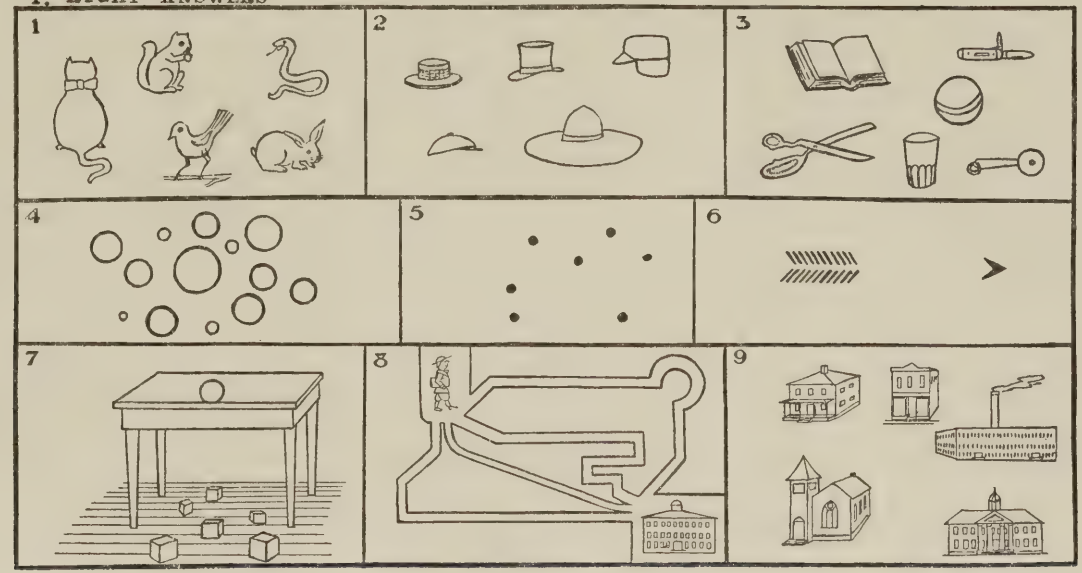


PLATE II
(*Reduced about two-fifths*)

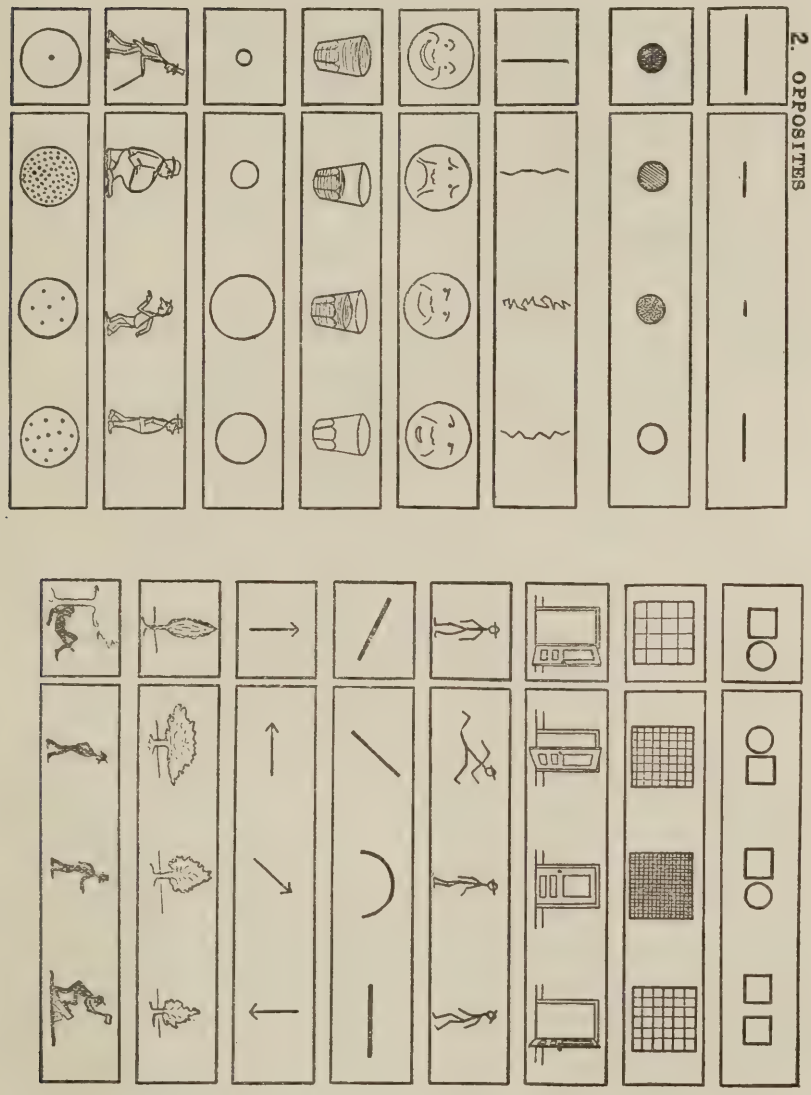


PLATE III
(*Reduced about two-fifths*)

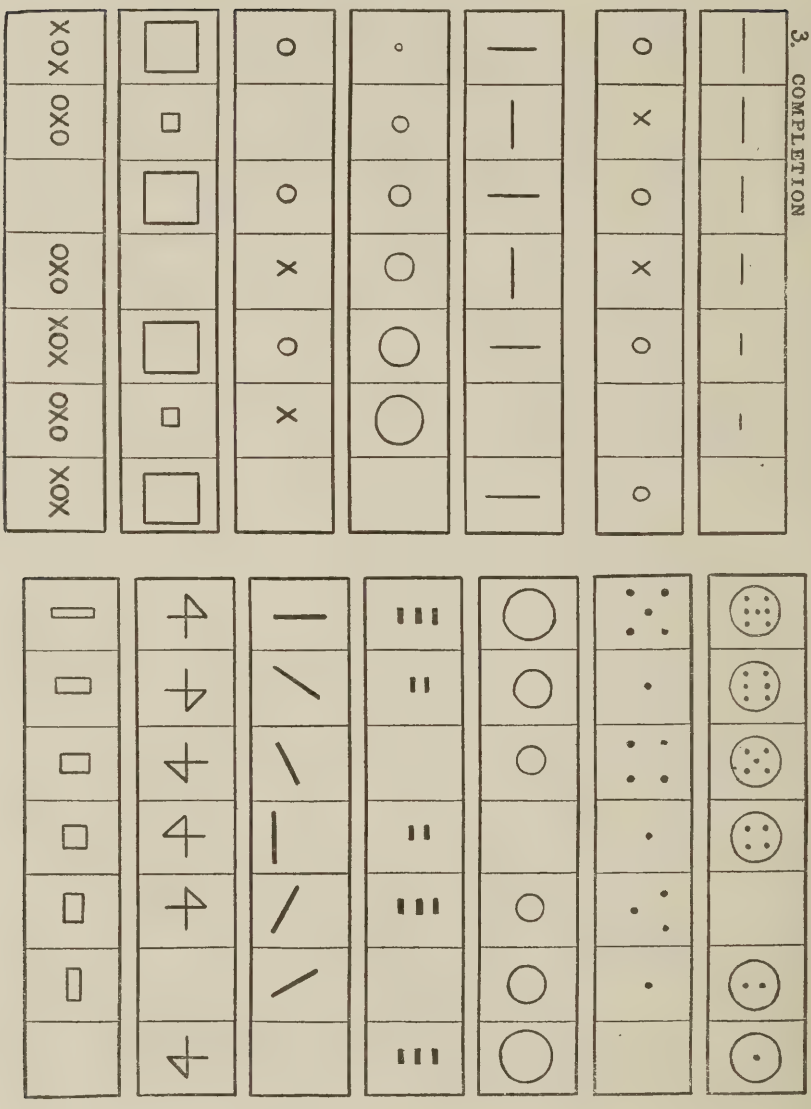
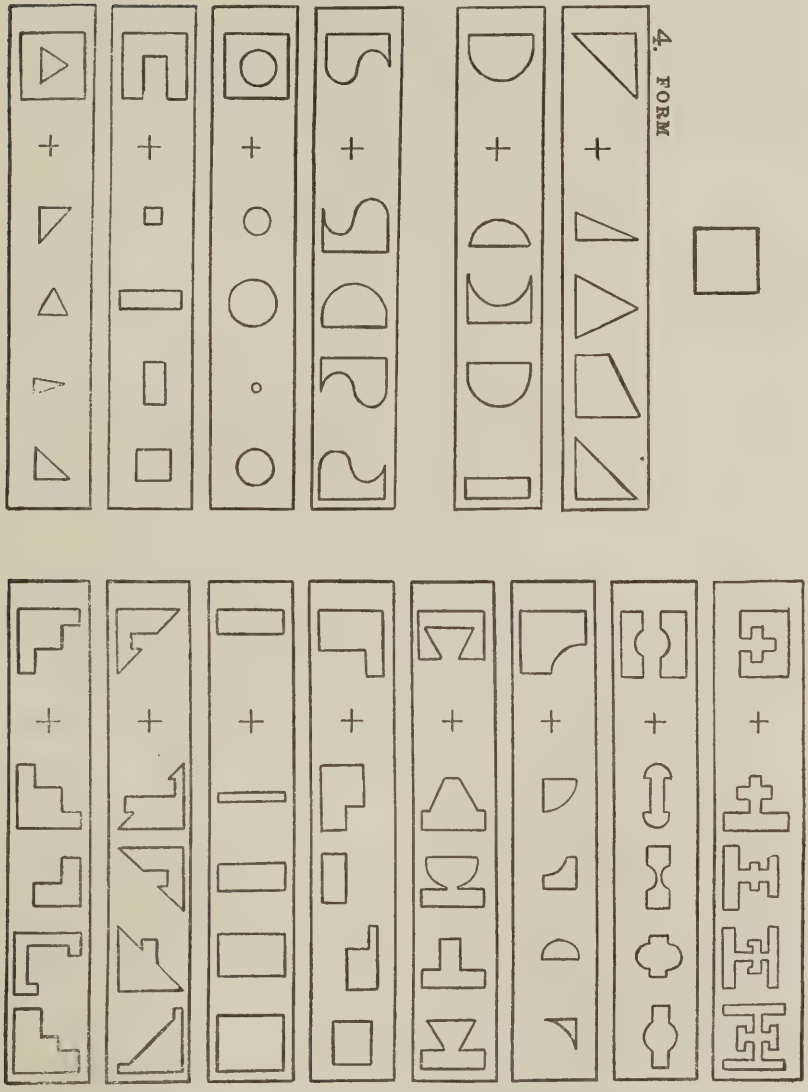
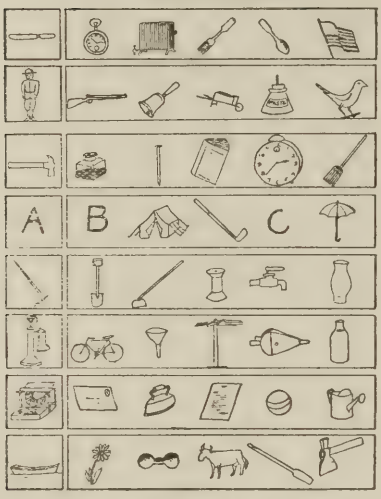


PLATE IV
(*Reduced about two-fifths*)



PLATES V AND VI
(Reduced about seven-tenths)

Plate V. ASSOCIATED OBJECTS TEST



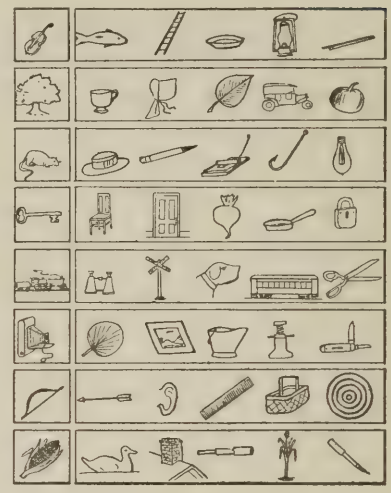
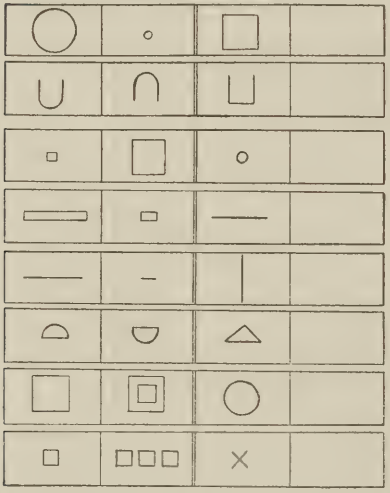
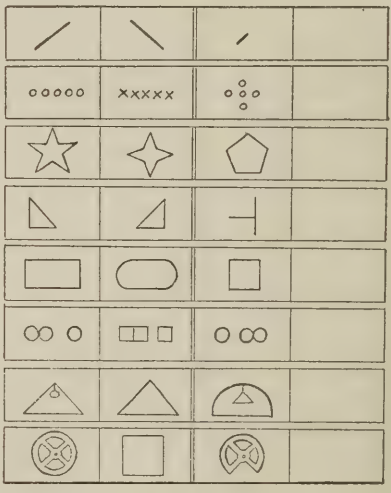


Plate VI. ANALOGIES TEST





first two, which were used for illustrative and demonstration purposes);

b. Series Completion test (fifteen parts);

c. Form test (twelve parts).

The second part included:

d. Opposites test (fourteen parts);

e. Right Answers test (sixteen parts);

f. Analogies test (fourteen parts).

These were bound in such a way that the two pages containing each test faced each other, making it unnecessary to turn a page during a test. The instructions for the various tests are given in the Appendix (II).

The time allowance was determined by asking the pupils to raise their hands when they had finished, and stopping the test when the first two or three had raised their hands. Since the computations of correlation with teachers' estimates were necessarily based on single rooms as units, and since it seemed desirable to ascertain what the younger pupils could do had they a longer time, the time allowance was not kept uniform throughout the preliminary test, but was increased somewhat in the lower grades. On the Right Answers test the time was kept uniform, varying from five to fifteen seconds for each part; on the others, the time varied from one minute to two and a half minutes on the various tests. In one 3A grade room the time allowance (which was made the same as in the 3B room where the test was first given) proved somewhat too long, so that about one-third of the pupils were able to finish each test before time was called. The result was that the slower pupils had an opportunity to work on through the test after the quicker children had finished, thus decreasing the average deviation of the scores in this room as compared with that in the other rooms, and partially obscuring real differences in ability.

The instructions for giving the test which had been perfected during the first tryout were worded to suit the new (group) conditions under which the test was to be given.

This preliminary form was then given as a group test in six rooms of the University Elementary School (hereinafter designated as "School U") and five rooms in the suburban town of Hinsdale

(designated "H"). In the former school the two parts were given with an interval of about three weeks between parts; in the latter, both parts were given on the same day, with an interval of about an hour and a half. About 250 children in all were given the test, but it proved necessary to discard some of the papers of pupils who were present for only one part of the test and whose records were therefore incomplete. Two hundred and thirty-four sets were obtained and used in computing the correlations. In seven of the eleven rooms the Stanford-Binet test had been given about three months earlier; in nine rooms the teachers' estimates of the pupils' intelligence were available. The schools, grades, number of pupils, and criteria available in each room are shown in Table I.

TABLE I

Schools in Which Preliminary Group Test Was Given, and Criteria Used in Computing Correlation-Coefficients

SCHOOL	GRADE	NO. OF PUPILS	MENTAL AGE	TEACHERS' ESTIMATES	TEACHERS' ESTS., REVISED
U	Kindergarten	18	*	*	*
	1A	10		*	*
	1A, 2B	12	*	*	*
	2A	22	*	*	*
	3B	13	*	*	*
	3A	21	*	*	*
H	South, 1	44	*		
	North, 1	14	*		
	S., 2	34		*	
	N., 2	14		*	
	S., 3	32		*	
		234			

The attitude of the pupils was throughout interested and active; the shortness of the time (about twenty minutes for each part) precluded fatigue or general loss of interest.

As the papers were scored, a full record was kept on each paper of both right and wrong responses and of parts omitted. In cases where two responses were required and only one was made, it was counted omitted; where two (or more) responses were made, one right and the other wrong, it was scored wrong. The detailed record of each child was tabulated on large co-ordinately-ruled sheets, one sheet to a room, to facilitate computation.

Arranging the Scale and Determining Its Validity

A variety of computations proved to be necessary in connection with the arrangement of the final form of the scale. Among the questions that had arisen in the course of the study were the following:

1. Should all parts of the Right Answers test be retained?

2. Should allowance be made, in scoring the Opposites and Form tests, for the right responses which would be made by chance alone?

3. Will a part of the tests (*e.g.*, three or four) show as significant results as all six taken together?

4. Which combination of tests is most significant?

5. Should the Right Answers test (or any other) be weighted?

6. In each test, what is the relative difficulty of the different parts?

7. Do these results correlate most closely with the mental ages (Stanford-Binet test), or with teachers' estimates? Why?

8. Do the various tests measure the same ability exercised in slightly different forms, or different abilities, or partly the same and partly different?

These questions, of course, can be answered only through statistical studies of the results, and most of them evidently call for computations of coefficients of correlation. Of the various methods of computing correlation, the so-called "Rank Method" of Spearman (using the formula, $\rho = 1 - \frac{6\,(\text{Sum } D^2)}{N\,(N^2 - 1)}$, where D is the difference between the rank of any individual in one series and his rank in the other, and N is the total number of individuals) was used in most cases, primarily because the teachers' estimates were usually arranged in ranked order, thus necessitating this method of computation. Furthermore, since the coefficients were used primarily for comparison among one another, and not in comparison with coefficients taken from some other investigation which might have been obtained by some other formula, the method which was necessary for some of the computations was used for

others as well. It may be remarked, however, that coefficients computed by this method vary less than .02 from those computed by the "product- moment" method. Where several rooms are included, as is usually the case, the coefficient given is the average (weighted by size of room) of coefficients for the various rooms, and the Probable Error is given.

We take up the questions suggested above in their order.

1. It became apparent before scoring had gone far that some of the Right Answers tests were so easy as to be worthless. Two such tests were eliminated before the scoring was finished, one because it was practically impossible to find cases which could be scored as absolutely right or absolutely wrong; the other, because its form suggested the proper response too readily. In order to determine what other parts should be eliminated, the 180 papers which had at that time been obtained were examined, and the right and wrong responses on each part tabulated (see Table 2). The number of errors made on each part of the test by the best quarter of each class, the second quarter, the third quarter, and the poorest quarter (or, in smaller classes, by the better half and the poorer half of the class) was then compared. Those parts of the test which revealed very few errors, or which did not show a fairly consistent decrease of errors from the bottom of the class to the top, were eliminated. In this way, five more tests were eliminated, leaving nine of the original sixteen parts.

TABLE II

Number of Right Responses Made on Each Part of the RIGHT ANSWERS *Test (Preliminary Group Form) by Better Halves and Poorer Halves of Seven Groups* (180 *Pupils*)

TEST NO.	DESCRIPTION	NUMBER RIGHT RESPONSES		PERCENT DIFFERENCE
		BETTER HALF	POORER HALF	
	(*Tests Retained*)			
2	Hats and Caps	71	58	.10
3	Plaything	64	53	.09
4	Circles	65	50	.13
5	Dots in Row	37	31	.09
6	Arrow	35	24	.19
7	Table, Ball, Blocks	78	66	.08
8	Four Paths	74	58	.12
9	Buildings	51	45	.06

(Tests Not Retained;
these are described, but not illustrated, in Appendix I)

10	Ball and Field	44	28	.22
11	Tools	82	75	.04
12	Toys	77	85	—.05
13	Animals to Eat	74	71	.02
14	Rectangle	Not Scored		
15	Heavy and Light Objects	Not Scored		
16	Animals	90	87	.02
17	Tools to Cut	81	77	.03

The retained tests correlated fairly consistently with the criteria of intelligence. One test which did show good correlation was eliminated, Terman's "Ball and Field" test (with a time allowance of fifteen seconds), which had been included in the group test for purposes of trial. This was eliminated for several reasons: (1) Some pupils were already familiar with it; (2) it was the only borrowed test in the lot; (3) it is more difficult to score satisfactorily than any of the others. The Right Answers test was therefore reduced to eight parts (in addition to the one used for demonstration) which it contains in the final form, illustrated in Plate I, and it is this final form which was used in computing the various coefficients of correlation.

2. In two tests, the Opposites and Form tests, there is a probability that a certain number will be marked right if the child simply marks at random or in some stereotyped form, and makes no mental effort whatever. There is also such a chance in the Associated Objects test, but since the requirement calls sometimes for one and sometimes for two pictures to be marked, the probability is very small that more than one or two will be correctly marked by mere chance. This sort of situation is usually met by subtracting a fraction of the errors from the rights in scoring, the fraction being such that a purely random response to the series of tests will result in a zero score. In the Form test the chance is one in four that any given part will be marked right, or a probability that if all twelve are marked, three will be right. This would be corrected by deducting one-third of the errors from the rights. Table 3 indicates that to make such correction raises the average coefficient of correlation with the mental age very slightly. The same is true with the

correlation with teachers' estimates. In at least two rooms such deduction has the effect of lowering instead of raising the correlation, probably because, as inspection of the papers shows, the less intelligent children tend to make fewer total responses in this test than do the brighter children, and thus to make a total of fewer errors (or at any rate, not many more) than the latter. But against this slight superiority of corrected scoring is to be weighed the considerable practical disadvantage of dealing with fractional scores (one-thirds and two-thirds), which retards and increases the probability of arithmetical error and inconvenience in scoring, listing, and computing, especially where large numbers of papers have to be handled rapidly. Hence, such a deduction has, in this study, been omitted.

In the Opposites test the probability of chance success is somewhat greater, one in three. The score would, therefore, be corrected by subtracting one-half the errors from the rights; then, if a child marks nine of the parts in a purely stereotyped or random way, he will probably mark three right and six wrong, which, by the formula, would give him a zero score. Although computations in several rooms show that the gain from such correction is not much greater than that in the Form test, the inconvenience of fractional scores can be avoided by the expedient of doubling the scores throughout the test, and this plan of scoring has therefore been adopted.

3. Computations showed that the sum of the scores of six tests gave, on the average, no higher correlations with the criteria of intelligence than do several combinations of four tests, in spite of the fact that the group is more significant than any single test. These coefficients are summarized in Table 3. Even if they had shown a higher correlation, it would have to be materially higher to compensate for the greater practical advantage of the smaller number, which is as many as can be given in the length of time it seems desirable to use for primary children, *i.e.*, from twenty to twenty-five minutes. But an inspection of the table of coefficients for the various combinations shows no such superiority. The significance of this fact will be discussed later, as we are at this point concerned primarily with the procedure followed.

TABLE III

Coefficients of Correlation (Rank Method) for Tests and Combinations

(ABBREVIATIONS USED: AO = Associated Objects test; SC = Series Completion; F = Form; Op = Opposites; RA = Right Answers; An = Analogies; w.dbl. = weighted double; R—1/2E = Number of Rights minus half the number of Errors.)

TEST OR COMBINATION	WITH MENTAL AGE		WITH TCHRS'. EST.		WITH REVISED ESTIMATES	
	Rho	P.E.	Rho	P.E.	Rho	P.E.
Associated Objects	.55	.06				
Series Completion	.58	.06				
Forms (Rights)	.50	.06				
Form (R—1/3E)	.52	.06				
Opposites (Rights)	.55	.06				
Opposites (R—1/2E)	.58	.06				
Right Answers (8 parts)	.52	.07				
Analogies	.38	.08				
Sum of six tests	.63	.04	.42	.04		
Sum of six tests (RA w.dbl.)	.64	.04	.40	.05		
SC + Op + RA + An	.69	.03	.45	.04		
SC + Op + RA (w.dbl.) + An	.69	.03	.40	.05		
SC + F + Op + RA	.69	.03	.41	.05		
Same, for School U, 5 rooms			.43	.06	.53	.06
AO + SC + Op + An	.68	.05				
AO + Op + RA + An	.67	.05				
AO + SC + Op + RA (w.dbl.)	.67	.04				
F + Op + RA + An	.66	.05				
SC + F + Op + An	.66	.05				
AO + SC + F + Op	.65	.05				
AO + SC + F + RA	.65	.05				
AO + SC + RA + An	.64	.05				
AO + F + Op + An	.63	.05				
AO + F + Op + RA	.63	.04				
SC + F + RA + An	.63	.04	.41	.09		
AO + F + RA + An	.63	.05				
AO + SC + F + An	.63	.05				
SC + F + Op	.66	.04				
SC + F + Op + RA (for same rooms as preceding group)	.71	.04				

4. Several of the fifteen possible combinations of four tests seem to have about equal value. In fact, the coefficients for any of the fifteen combinations indicate considerable value. Two combinations, however, stand out somewhat above the others, each showing an average correlation coefficient of .69 (±.03), the "r" equivalent being about .70 or .71. These two combinations which

show the highest correlation with the other criteria of intelligence are:

Right Answers test	Right Answers test
Opposites test	Opposites test
Series Completion test	Series Completion test
Form test	Analogies test.

Several other combinations of four seem to be almost equally worthy of consideration. The two combinations described, however, yield almost identical correlation coefficients throughout the rooms, with both teachers' estimates and mental ages. Practical considerations finally led to the choice of the group including the Form test, as being easier for untrained examiners to administer than the Analogies test. It is rather surprising to notice that two tests which individually show rather marked difference in their degree of correlation with intelligence criteria should, in combination, yield such similar results, illustrating rather strikingly the value of a combination of tests in order that they may supplement one another. The two tests seem to measure a number of identical elementary processes, form perception, comparison, etc., although there are undoubtedly marked differences in other processes which they measure. The final form of the scale, therefore, was made up of the first combination of four named above.

Another fact, familiar to the statistician, but too often overlooked by the layman, has a bearing on the significance of these coefficients, viz: the heterogeneity or homogeneity of the data from which they were computed. Not infrequently do we find cited coefficients of correlation between test results and various criteria of validity ranging from + .75 to + .85 or even higher; but on examination we often find that the cases cover a wide range of ages or grades, the effect of this heterogeneity of data being, of course, to magnify whatever correlation there may be and to raise the coefficient. The coefficients here cited, on the other hand, are averages of coefficients computed separately for highly homogeneous groups, i. e., single grades or rooms. Had data from the three or four grades been assembled and a single coefficient computed for this relatively wide range of ability (which was not possible, ow-

ing to the varying time allowances in the experimental testing) it would undoubtedly have been materially higher.

5. The question of weighting the tests then arose, particularly with reference to the Right Answers test, which includes only eight parts, as compared with the twelve or fourteen parts in each of the others. In a number of rooms computations were made, both without weighting the test, and with the score on this test doubled. The comparative results summarized in Table III showed that, while the significance of the Right Answers test varied from room to room within a limited range, in the long run weighting the test made no significant difference, one way or the other. The question was therefore settled on purely practical grounds, in a way which will be described later, but which, in effect, amounted to a weighting of one and one-half.

6. The question of relative difficulty of parts within each test was settled by taking about 100 papers and counting the rights and errors on each part. The object of this was to arrange the final scale with the easier parts at the beginning of each test and the harder parts at the end of each test. One reason for such arrangement is that the pupil may have the encouragement to continue his efforts which comes from attacking the more easily solved problems first. Such an arrangement, moreover, makes the final score not merely an index of his speed, but a test of his ability to solve problems of constantly increasing difficulty; so that the child who makes a high score is not only a more rapid thinker and worker, but has greater insight and better judgment. The result of this investigation is, so far as the four tests of the final scale are concerned, revealed in the order of the parts within each test as shown in Plates I to IV. The order of difficulty in the other two tests (Plates V and VI) is not so accurately measured.

7. An inspection of the coefficients in Table 3 shows that the tests correlate considerably more highly with the Stanford-Binet mental ages than with the teachers' estimates. The test scores on the combination of tests chosen for the final scale showed correlation coefficients with mental ages ranging from .51 to .87, and averaging (average weighted by number of pupils in room) .69+, which is the equivalent of $r = .71$ by the product-moment method

of computation. With the teachers' estimates, on the other hand, the coefficients ranged from +.01 to .74, averaging .41. This average was materially reduced, of course, by the estimates for the room which showed the lowest correlation. It proved impossible to check up the teacher's estimates with the mental ages in this room, since for some unknown reason only a few of the Binet records for this group could be found. That some very different criterion was used in making these estimates seems evident, however, from the few Binet figures which were available for children of this group; thus, one child, listed eighth in the group of ten, but highest in the group test, showed an intelligence quotient of 142. Eliminating this room, the average coefficient of correlation between the group test and the teachers' estimates is .43. The revised estimate, described later, showed a markedly higher agreement, both for this room and for the entire school. In a later section the significance of the difference between correlation-coefficients computed with the mental age and with the teachers' estimates will be discussed at greater length.

8. The make-up of the final form of the scale has already been indicated. The mechanical arrangement proved something of a problem, due to the form limitations which the nature of three of the four tests imposed. It was finally met as shown in Plates I to IV.

The scoring system adopted was based on 100 as a perfect score, which is made up as follows:

Right Answers test, 8 parts, rights × 3, perfect score	=	24
Opposites test, 14 parts, (rights—½wrongs) × 2, perfect score	=	28
Series Completion test, 12 parts, rights × 2, perfect score	=	24
Form test, 12 parts, rights × 2, perfect score	=	24
Total	=	100

The scoring could, of course, have been arranged on a scale of 50 instead of 100, by not weighting the last three tests, and weighting the Right Answers test by 1½, as described earlier; but the present plan was adopted to attain the purely practical end of avoiding fractional scores, which would otherwise appear in the Right Answers and Opposites tests.

Following the determination of the final make-up of the scale,

drawings were prepared, electrotypes made, and 1500 test blanks, of the form shown in the insert (Plates I to IV), were printed. In this form the scale was used during the month of May, 1920, in testing over 1300 pupils in the schools listed on pages 22 and 23, for the purpose of obtaining norms.

In the first thirty rooms listed, the testing was done by the author personally (except for about fifteen pupils in System E, who were tested a few days later by an experienced primary teacher). In the other schools it was done by superintendent or teachers. Blanks were sent to these schools with the definite object of ascertaining whether the results obtained by other examiners were comparable with those obtained by the author, and thus to secure a check on the practical reliability of the scale and its accompanying instructions. In most of these cases, a careful examination of the test papers seemed to prove that the test had been given with as much care as by the author, and the results were fully comparable. In two first grade rooms (System O) the time allowance had evidently been made much too long to permit the scores to be compared with those from other schools. Returns from these rooms have not, therefore, been included in the summary of scores. In two other rooms (System N) there was some evidence that the directions on one or another of the Right Answers tests had not been accurately given, so that an indeterminate portion of the pupils in those two rooms scored three points lower than they should. In spite of this fact it was deemed best not to throw out these two sets of papers, since their influence on the norms is small. The instructions therefore seem to be adequate for anyone who exercises reasonable care in following them. The paragraph (see Appendix II) emphasizing care in observing the time factor was inserted after the scores herein described had been obtained, with a view to preventing a recurrence of the oversight described above.

It was found by the author that to give the test required from about twenty minutes (in third grade rooms) to twenty-five minutes or more (in first grade rooms), the variation being due solely to the amount of time required for "preliminaries," inserting name, age, grade, etc. After this had been done, the actual time for work on the testing was uniformly about eighteen minutes.

It was the original intention to give the test in only the first three grades. At one superintendent's request, however, the author consented to give it in one higher room, characterized as "slow," and including fourth and fifth grade pupils. The results were so directly comparable with those of other grades that other fourth grade rooms were tested, partly in order to get fourth grade norms, and partly to get scores from those nine-year-old children who are above the third grade. No kindergarten children were included in this final series of tests, hence no kindergarten norms, and no age norms below six years have been established, although the experience with the preliminary group form on which the correlations were computed proved that the test is adapted to many kindergarten children.

Distribution of Scores

The scoring was found to be an easy task. After practice the author found it possible to score as many as eighty papers per hour. Instructions for scoring are given in Appendix III. The scores for each pupil, together with age, sex, and grade, were listed in ranked order, and sent to the teachers with the request that they make corrections and return to the author.

The distribution of scores made on the final form of the scale is shown in Tables 4, 5, 6, and 7, in graphic form in Figures 1 to 5. The schools and communities represented may be characterized as follows:

1. Evanston, Ill., South Side Schools, District No. 76 (hereinafter designated as "System E"), 22 rooms. A residential and university suburb, selected partly because it represents a considerable range of social strata, and is therefore fairly representative of an American middle-class community.

Four schools are included, as follows:

School A. Five rooms. In western part of suburb; middle-class neighborhood, with many laborers' homes. In first grade, thirty-five per cent of the children bear Polish names; in lesser degree this is characteristic of other grades. Some other European nationalities are represented, although American ancestry predominates.

School B. Five rooms. Somewhat better neighborhood than the former; fewer foreign names, but largely laborers' homes.

School C. Six rooms. Moderately well-to-do neighborhood; many business and professional men.

School D. Six rooms. In best part of the town; many wealthy and cultured families represented.

Scores by age-groups are given for the system as a whole in Tables IV and V; grade-scores are shown in Table 6 for the separate schools as well as for the entire system.

2. Jackson School, Chicago. Designated as "School J." Eight rooms. In the heart of the Italian district, three blocks from Hull House. Ninety-eight per cent of the children are of Italian parentage. Age and grade scores given.

3. New Hampton, Iowa. Designated as "System N." Four rooms. County-seat town of 2500 population; agricultural community. Age and grade scores given.

4. Lincoln School of Teachers' College, New York. Designated as "School L." Three rooms. An experimental school in a wealthy neighborhood; distinctly a selected group. Ages were not obtainable here, so only grade-scores (grades 1, 2, and 3) are given.

5. Osage, Iowa. Designated as System O." Four rooms. County-seat town of 3000 population; agricultural community; much like System N, but probably somewhat higher cultural level. Scores for two first-grade rooms had to be discarded, as explained earlier, and no fourth grade tested; hence, age-scores for remaining (second and third) grades are not summarized, since they would not be fully representative.

TABLE IV

Distribution of Scores by Ages and Sexes (Grades 1 to 4)

SCHOOL	SEX	AGE LAST BIRTHDAY	NUMBER OF CASES	SCORES			
				HIGH	FIRST QUARTILE	MEDIAN	THIRD QUARTILE	LOW
E.	Boys	6 years	71	65	34	22	10	0
		7 "	74	71	45	38	25	0
		8 "	88	93	64	48	34	7
		9 "	64	87	69	59	41	9
		*10 "	34	86	75	59	43	14
		*11 "	26	81	75	58	39	31
		*12+ "	7	78	..	37	..	14
			364					

E.	Girls	6 years	60	54	31	23	15	0
		7 "	69	74	46	37	20	0
		8 "	71	78	60	48	28	4
		9 "	54	89	65	51	39	14
		*10 "	29	81	69	52	40	27
		*11 "	7	79	..	63	..	46
		12+ "	6	50	..	30	..	20
			296					
J.	Boys	6 years	30	23	9	6	1	0
		7 "	24	63	22	17	1	0
		8 "	37	69	26	19	10	3
		9 "	30	68	34	26	19	9
		*10 "	24	66	39	34	19	12
		*11 "	11	68	44	31	20	13
		*12+ "	9	60	..	39	..	21
			165					
J.	Girls	6 years	36	33	7	4	2	0
		7 "	29	67	17	11	5	0
		8 "	36	46	27	14	7	0
		9 "	25	46	35	21	14	2
		*10 "	22	42	33	22	15	6
		*11 "	6	50	..	33	..	23
		*12+ "	5	74	..	30	..	22
			159					

* In this and later tables it should be borne in mind that the scores given for ages 10 and over are not representative of those age-groups, and are not reliable as norms, since they include few or no pupils above grade 4. Those whose scores are shown here are, therefore, to a considerable extent over-age and retarded children; hence their scores average lower than the normal score would be for those ages.

TABLE V

Distribution of Scores by Ages (Both Sexes) Grades 1 to 4

SCHOOL	AGE LAST BIRTHDAY	NUMBER OF CASES	SCORES HIGH	FIRST QUARTILE	MEDIAN	THIRD QUARTILE	LOW
E.	6 years	131	65	33	22	12	0
	7 "	143	74	45	37	23	0
	8 "	159	93	62	48	33	4
	9 "	118	89	67	56	41	9
	*10 "	63	86	70	55	41	14
	*11 "	33	81	73	59	42	31
	*12+ "	13	78	..	50	..	14
		660					

J.	6 years	66	33	9	5	2	0
	7 "	53	67	19	13	6	0
	8 "	73	69	26	16	9	0
	9 "	55	68	35	25	17	2
	*10 "	46	66	36	29	18	6
	*11 "	17	68	44	31	26	6
	*12+ "	14	74	..	37	..	21
		324					
N.	5 years	9	41	23	12	7	5
	6 "	16	47	32	24	13	7
	7 "	32	75	47	34	14	5
	8 "	21	65	56	39	30	22
	9 "	19	77	65	49	34	12
	*10 "	8	76	62	45	31	25
	*11 "	4	46	..	30	..	11
	*12+ "	1	..	..	53	..	..
		110					

TABLE VI

Distribution of Scores by Grades, Both Sexes

SCHOOL AND GRADE	NUMBER OF CASES	SCORES HIGH	FIRST QUARTILE	MEDIAN	THIRD QUARTILE	LOW
1st, E, School A	49	50	27	15	5	0
" B	23	62	30	17	11	3
" C	54	57	30	24	15	3
" D	48	65	33	25	14	0
2nd, E, School A	29	68	37	28	25	7
" B	22	50	43	28	18	9
" C	37	77	51	41	32	18
" D	75	69	49	42	32	15
3rd, E, School A	29	73	60	47	38	23
" B	28	74	60	45	32	21
" C	52	85	64	55	40	20
" D	59	93	67	56	48	27
4th, E, School A	40	80	66	51	37	14
" B	37	87	73	56	48	27
" C	32	78	66	60	55	31
(4B and 4C only) D	22	89	76	68	54	21
5th, E, School A	13	80	76	61	44	35
(5B and 5C only) B	18	81	77	68	52	35
System E, Gr. 1	174	65	30	21	11	0
2	163	77	47	35	26	7
(Four 3	168	93	64	52	40	20
Schools) 4	131	89	71	59	48	14
*(5)	(31)	(81)	(77)	(65)	(54)	(35)
	667					

* These scores not representative of entire 5th grade or of entire city, hence not reliable as norms.

System N,	1	28	51	26	16	9	5
	2	32	75	42	33	24	9
	3	34	68	56	46	32	12
	4	16	77	68	54	40	18
		110					
School J,	1	91	36	9	4	2	0
	2	69	35	19	12	6	0
	3	107	68	32	22	14	3
	4	76	74	39	32	22	4
		343					
School L,	1	16	62	47	38	28	15
	2	17	74	68	61	50	35
	3	18	92	79	75	62	42
		51					
System O,	2	33	78	59	43	30	9
	3	37	85	64	53	38	22
		70					

TABLE VII

Decile Distribution of Scores by Ages (E, four schools, 660 pupils)

	6 Y'RS (131)	7 Y'RS (143)	8 Y'RS (159)	9 Y'RS (118)	10 Y'RS (63)	11 Y'RS (33)	12+Y'RS (13)
Highest Score	65	74	93	89	86	(81)	(78)
First Decile	45	53	71	75	80		
Second "	36	47	65	70	75	(73=Q)	
Third "	30	44	60	65	67		
Fourth "	27	42	53	60	63		
Fifth (MEDIAN)	22	37	48	56	55	(59)	(50)
Sixth Decile	18	31	41	50	51		
Seventh "	15	25	35	42	43	(42=Q)	
Eighth "	8	19	27	35	39		
Ninth "	5	12	21	24	32		
Lowest Score	0	0	4	9	14	(31)	(14)

FIG.1 (Table 6). Distribution of Scores by Grades. (System E, 4 Buildings, 636 Pupils)

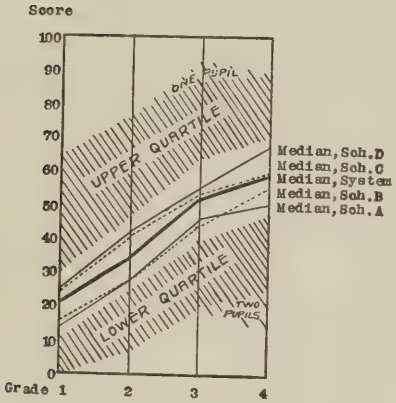


FIG.2 (Table 6). Median Scores by Grades, Five Communities.

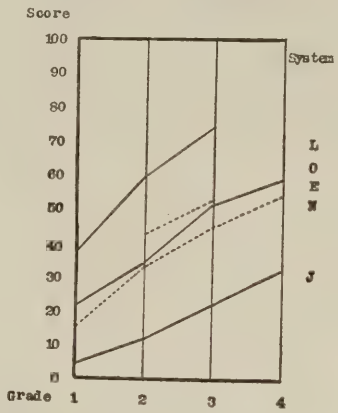


FIG.4. Median Scores by Age and Grade, Three Systems.

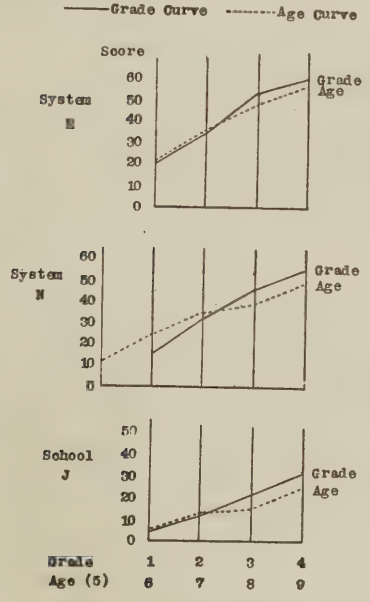


FIG.3 (Tables 4,5). Median Scores by Age and Sex.
(System E, 660 Pupils)
(School J, 324 Pupils)

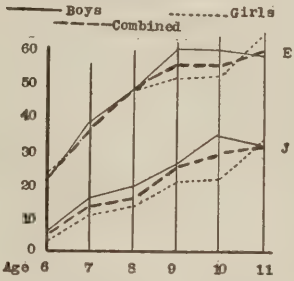
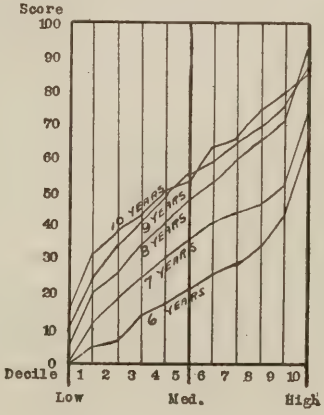


FIG.5 (Table 7). Decile Distribution of Scores by Ages (System E).



Analysis of Score Distributions

An examination of the tables and graphs in which the results of the tests are summarized reveals a number of interesting facts.

The curve which traces the median (or either quartile) score from grade to grade, or from year to year, is normally a straight line. That is to say, the child makes substantially equal progress in ability each year during the three year-intervals represented. This is true, regardless of the absolute score, being characteristic of the median and quartile curves in schools of widely varying status. Two qualifications are important in this connection, although neither affects the validity of the conclusion stated.

First, there is apparent in some of the better fourth grades, notably E and N (and it would probably be equally manifest in O and L, had we fourth grade data for these schools), the slowing-down effect due to the approach to the maximum possible score, 100. This effect is particularly noticeable in the first quartile of the system E distribution. This indicates that, while 100 is theoretically possible of attainment, and probably would be attained by some were the test given in higher grades, the difficulty of the hardest tests prevents most scores from going above 90. Only two of the 1242 papers here summarized have reached that mark. That this is due to the nature of the scale, and not to this particular stage of growth, seems apparent from the absence of this bend in the curve of the J school, whose scores stand well below the influence of these hardest tests.

It should be remembered in this connection that the scale was designed for the first three grades, and that its use in the fourth grade was in the nature of an afterthought. In the first three grades the influence of this approach to the maximum is not observable save in the upper part of the first quartile. It is of interest to find that the scale lends itself so readily to use in the fourth grade, thereby making possible comparative studies with scales designed for intermediate and upper grades and involving the use of reading material.

A second qualifying comment refers to the age distributions. It will be observed that median scores for ten, eleven, and twelve (and over) year old children not only do not ordinarily rise much

above the nine-year median, but frequently are lower. This is due, of course, to the fact that (with a very few exceptions in two E rooms which included some fifth grade children) the only pupils of these ages who were tested were in the fourth grade or below, hence over-age and presumably averaging below normal intelligence. The median and quartile scores for these years are not, therefore, valid as age-norms, and have therefore been distinctively marked in several tables, with parenthesis or asterisk.

That the form of distribution of these scores resembles the normal probability curve is apparent both from the quartile distribution of Tables IV, V, and VI (Fig. 1) and the decile distribution of Table VII (Fig. 5). The middle half of the scores have, on the average, a smaller range than either the upper or lower quartiles, indicating the "bunching" of scores around the median, and their "scatter" at the upper and lower extremes. This was observed to be characteristic, in less degree, of the score sheets for the various rooms, and is, indirectly, a confirmation of the reliability of the scale as a measuring instrument.

Table IV (Fig. 3) shows the sex differences in the E system and the J school. In the former, they are so small as to be negligible, the only deviation (ages 9 and 10) being partly accounted for by the limited number of cases which, scattered over a considerable range of scores, may displace the median. The use of the average instead of the median reduces the difference at age 9 by two points. In the J school, however, there is a constant sex difference, the boys scoring about five points higher than the girls at each age. The cause of this difference has not yet been determined; possibly a more aggressive adaptability to strange social customs may have led the boys in this Italian neighborhood to do better than the girls; but this is purely a speculative suggestion.

The most striking differences revealed are those between median and other representative scores in different schools (Table IV). Between the four E schools there are marked differences, corresponding definitely to the differences in social status of each neighborhood described on pages 22-23. The two small-town systems do not vary far from the E norm; the lower standing of N is partly due to the slight error, described earlier, in giving the test, which

makes several of the scores three points lower than they should be; the advantage is, however, with the town (O) which prides itself on its school system and cultural atmosphere. The L school and the J school deviate widely from the E norm. In each case the direction and relative amount of deviation is just what our knowledge of the two communities would lead us to expect, in the light of the many studies that have been reported on the relation between intelligence and social status. The scores seem to warrant the selection of E as a fairly representative American community, as well as the characterization of the L school as a highly selected group, and the J school as representative of a community distinctly below the standard. We have no data to warrant entering into a discussion of the cause of these variations, whether they be racial, environmental, or physiological (nutritive, sensory, etc.). It is possible that significant supplementary data may be available from these schools at some future time to aid in such analysis.

It is of interest to note the relation between the age scores and grade scores in the three schools from which we have age data, E,J, and N. This has been put in graphic form in Fig. 4. In each case, the 6-year age median is slightly above the first grade median, and the 7-year age median above the second grade median; while the 8-year age median invariably falls below the third grade median, remaining below in the ninth year and fourth grade. The same type of phenomenon is in general characteristic of the first and third quartiles in all three schools, except that the age quartile ordinarily falls below the grade quartile a year earlier (*i.e.*, 7 years, second grade); the first quartile, system N, offers the single exception. That this phenomenon should occur uniformly in three so widely different systems, totaling over a thousand pupils, indicates that it is not a chance occurrence, but a definite demonstration, by the group scale, of certain normal conditions. These conditions may be crudely summarized as follows:

The average	six-year old	is brighter	than	the	average	first-grader
"	"	seven-year old	"	"	"	"	"	second-grader
"	"	third-grader	"	"	"	"	"	eight-year old
"	"	fourth-grader	"	"	"	"	"	nine-year old

The key to these differences is to be found in the complex make-

up of each grade. A given grade contains a variety of elements,—bright, normal, and dull normal-age children, over-age children of normal or inferior intelligence, under-age children of normal or superior intelligence, and possibly even under-age dull children or over-age bright ones. The dullness and brightness, moreover, occur in widely varying degrees. The median score will consequently be a function of the relative weight of these diverse factors. Table VIII presents an analysis of system E (668 pupils).

TABLE VIII

Median Intelligence Scores of the Various Age-Grade Groups (System E)

		Grade 1 Median 21		Grade 2 Median 35		Grade 3 Median 52		Grade 4 Median 59		Grade 5 Median 65	
Age	Age Median	Med.	No.	Med.	No.	Med.	No.	Med.	No.	Med.	No.
6	22	21	116	41	13	40	1				
7	37	24	47	42	90	52	13				
8	48	11	6	32	48	53	95	63	11		
9	56	11	2	32	13	51	42	60	57	76	2
10	(55)	14	3	30	2	51	7	55	40	67	11
11	(59)					36	6	60	16	60	11
12						20	1	29	4	61	4
13						38	2	29	2	50	3

If the progress of every pupil through the grades were uniform and regular, or if acceleration of superior pupils were as frequent as retardation of backward pupils, the age-median and grade-median would coincide. But as every teacher knows, and as this and every age-grade table reveals, acceleration is (largely because of the administrative difficulties it involves) much less frequent than retardation. Hence the middle and upper grades tend to become selected groups, including more bright children of normal age who have not been accelerated than dull children of normal age who have been left scattered through the lower grades. The grade-median will tend, therefore, to rise gradually above the age-median in successive grades. Let us consider the composition of a typical grade-group.

An analysis of the fourth grade shows that it consists in the main of 9-year olds and 10-year olds. The 9-year olds include bright, normal, and perhaps some dull pupils, but evidently fewer dull than bright, since there are so many dull 9-year olds in lower grades, and since the remainder average of superior 9-year old

intelligence (60:56). The 10-year olds are, as a rule, normal or dull, but not sufficiently dull to bring the intelligence-median for the group (55) far below the grade-median (59), and hardly below the 9-year age-median (56). Besides these, we find eleven accelerated 8-year olds, whose presence raises the grade average. We find also sixteen 11-year olds, enough of whom are over-age for apparently other reasons than dullness to keep the intelligence-median up. Finally, we find a very few very dull, over-age pupils. The large majority of distinctly dull 9-year olds and older children, who would otherwise bring down the grade-median, are found scattered back through the third, second, and first grades with younger children whose average intelligence is not so far different from theirs. Thus, because of the preponderance of bright children over dull, the grade-median is kept above the corresponding age-median.

In the first grade, however, and to a certain extent in the second, we find an important modifying factor. The first grade includes not only those normal-age pupils whose ability warrants their classification there, but all those dull normal-age and over-age children who average, as the table indicates, so much lower in intelligence that they would be found in a lower grade, were there any lower. The consequent accumulation of mediocre pupils, notably the 8-, 9-, and 10-year olds more than offsets the small group of non-accelerated bright pupils of normal age. This condition has not yet been completely overcome in the second grade, as the number of retardates and their average intelligence shows. Hence the first two grade-medians fall slightly below the corresponding age-medians.

The age-median curve and the grade-median curve, therefore, must cross at some point shortly after the second grade. By the time the third grade has been reached, retardation has produced the distinctly selective effect on the make-up of the grade which has been described. The fact that this normal, although rather fine-drawn over-balancing is reflected so definitely by the group-scale in three schools of so widely differing character, seems to be a verification of both its validity and reliability.

Norms

Were we required to give definite age-standards for this scale, such standards would of necessity, at this stage, have to be tentative, and based on some definite assumption concerning the "average American community." As a starting point the writer would, although with some hesitation, probably select an approximation to system E (for the reasons hereinbefore mentioned) as a basis, and suggest as tentative median scores the following:

Age 6 years (last birthday), median score 22
" 7 " " " " " 34
" 8 " " " " " 46
" 9 " " " " " 56

The last named figure reflects the approach to the maximum, described earlier.

For the various grades (keeping in mind that the scores were obtained in the last school month of the year, although in systems practicing term or mid-year promotion), the following may be suggested:

Grade 1, median score 21
" 2, " " 34
" 3, " " 47
" 4, " " 58

These norms are, of course, merely tentative suggestions and will be revised or confirmed as added data accumulate.[1] The var-

[1] Additional records, comprising 1098 additional cases tested during the second half of the school year, have been collected by the Bureau of Educational Research since the publication of the scale. These, together with the results herein presented (except those from School J which were not included for reasons previously mentioned), yield the following norms, representing about two thousand cases:

Grade 1, Median Score 19
" 2, " " 35
" 3, " " 48
" 4, " " 59

The net result of the addition of these later data is to lower three points the first-grade norm and raise those for the second, third, and fourth grades one point each. It is regrettable that absence of data concerning these schools does not permit further analysis of the educational and social character of these groups, nor presentation of revised age-norms. It should be remembered in using these norms that tests made early in the school year may be expected to yield scores considerably lower.

ious modifying factors that have been mentioned should, of course, be kept in mind in using the figures given.

Teachers' Estimates

It is apparent from an inspection of Table 3 that the results of this test correlate considerably more highly with the Binet-Stanford mental ages than with the teachers' estimates of intelligence. With the Binet scores, the combination of tests used in the final scale gives an average coefficient of correlation (rho) of .69, while with the teachers' estimates the average coefficient is about .41.

This lower correlation means that the teachers, in making their estimates of the pupils' intelligence, fail to take into sufficient account certain factors which the two tests (Binet and group test) measure, or else that they take into account certain factors which neither of the tests measure. What some of these factors may possibly be will occur at once to every reader. Such considerations as irregularity of attendance, physical or sensory defects, lack of interest in school work, some unfortunate attitude toward teacher or school, imperfect mastery of the language, laziness, etc, which affect unfavorably the quality of school work, may readily be imagined as leading the teacher to underestimate the child's real ability, which a test score would probably not thus discount; while such factors as pronounced interest in work, high responsiveness to the teacher's wishes (especially if the child's intelligence be not low but moderate), special aptitude for reading or number work, perseverance, commendable social attitude, an intelligent facial expression, or even neat personal appearance, might conceivably lead a teacher to overestimate the child's native ability. These factors, of course, an intelligence test measures not at all, or to a very limited degree.

On the other hand, it should be remembered that tests have, by their very nature, certain limitations as measures of intelligence. Temporary indisposition or disturbance may cause a child to do more poorly on a single test than his real ability warrants, while the reverse can seldom occur. Under the conditions of a group test, even more than in an individual test, such deviations are to be expected, since the personal rapport between examiner and child which will lead the child to do his best, can hardly be established;

and a chance distraction may lower a child's score disproportionately if it comes at a critical point and there is no opportunity for repetition of instructions. It should be borne in mind, however, that insofar as such cases have occurred, they have affected the correlation coefficients between the group scale and the Binet test scores, as well as between the test scores and the teachers' estimates; so that if one considers the former coefficients as reasonably high, it would indicate that the injustice worked on the pupils by the conditions of the group method cannot be excessive. Furthermore, let us not forget that no one yet knows what tests or scales will measure intelligence accurately; and it would be dogmatism to assert that it is only the teachers' estimate, and not the scale, which needs to be improved. The practical bearings of these limitations of the testing method are briefly discussed at a later point.

Nevertheless, the non-intelligence factors previously mentioned are, it would seem, responsible at least in part for the lower correlations of the test scores with teachers' estimates than with mental age figures. In order to raise the coefficient of correlation with teachers' estimates, such estimates should be guided. That is, it should be made clear to the teacher (as far as our limited information is possible) what is meant by "general intelligence"; she should be reminded what factors to take into account and what factors to neglect, and reminded that the child's school marks, while significant, are not an accurate measure of his bare native ability. She should be asked, moreover, to make her estimates not in such vague, undefined terms as "good," "fair," etc., but in terms of some quantitatively defined scale. Several such scales have been suggested in educational literature, varying from a mere ranking of the pupils in order with reference to a certain list of qualities (such as the one described later in connection with the re-estimates in School U) to elaborately worked out attempts to define the characteristics of ten or more graded intelligence groups. The scale used in rating army officers and various plans based thereupon,[2] wherein the estimator makes his own scale, placing five

[2] For example, that of Rugg, H. O., "Rating Scales for Pupils' Dynamic Qualities; Standardizing Methods of Judging Human Character," *School Review*, XXVIII: 337-349, May 1920.

carefully chosen individuals at five numbered points (low extreme, low, average, superior, high extreme), and rates the members of the group numerically by comparison with these "standard" individuals, has much to commend it. The writer used in certain cases a somewhat similar scale, with the following written instructions, sent to the teachers from whom estimates were to be obtained:

"Please give your estimate of the pupils' general intelligence, using the following numerical symbols:

1. Exceptionally superior intelligence
2. Above average
3. Normal or average intelligence (for the group)
4. Below average
5. Distinctly deficient.

By 'General Intelligence' is meant not merely the pupil's standing in school subjects, but his ability to direct and hold his attention to tasks, to adjust himself to new sorts of situations, and to be critical of his own efforts; both in school situations and outside of school" (essentially Binet's threefold conception of general intelligence).

Almost any definite plan for guiding teachers' estimates of intelligence is better than no plan. Several lines of evidence arising in the present study have confirmed this belief.

1. In several cases, teachers have given verbal comments on individual pupils which indicate that their own recorded estimates of those pupils are based on other than ability factors. A single typical example will suffice. One girl, ranked in the "unsatisfactory" group by her teacher, was described to the writer by that teacher as "bright, but lazy"; her Binet intelligence quotient was 108, and in the group test she stood fourth among thirty-two (percentile rank 86.5). A number of similar instances occurred in other rooms.

2. In one room (H, N. 2) the teacher had not made a list of estimates before the test was given. After it was given, but before the papers had been scored, the writer sent to the teacher a list of the pupils who had taken the test and asked that she rank them, taking occasion to explain the meaning of "General Intelligence," in the terms quoted above. The resulting estimates from this group

showed a higher correlation (.63) with the test scores than those of most other lists of estimates, and are comparable with the coefficients obtained with the mental ages.

3. The principal of school U, convinced that many of the original (unguided) lists of teachers' estimates were based largely on considerations other than general intelligence, asked his teachers, some four months after the first estimates were made, to re-estimate their pupils, taking into account the following definite factors: Age Factor, Attitude toward school work, Quality of school work, Self-confidence of pupil, Timidity of pupil, Resourcefulness, Memory, Ability to work, Ability to discriminate or judge, Social qualities. The revised estimates made under the guidance of these instructions revealed in practically all cases a higher correlation, both with the mental ages and with the group test scores, than the earlier list of estimates. Thus, where the average correlation between the group test scores and the original estimates was .43, with the revised estimates it rose to .53. It is true, of course, that the teachers had access to both sets of test results when they made their revised estimates, and the objection might be offered that the revised estimates represented not the independent judgment of the pupils' degree of possession of the traits specified, but an effort to approximate the test scores. A critical comparison of the various lists, and personal acquaintance with the teachers, however, leads the writer to believe that there was no conscious attempt on the part of any teacher to make her estimates fit the scores, but that the revised estimates are just what they profess to be.

4. Still another bit of evidence bearing on teachers' estimates was collected as a result of the request made that teachers go over the list of scores sent them, correct ages, grades, and names, and make comments on exceptional pupils, or pupils whose scores differed materially from what the teacher judged their relative intelligence to merit. The teachers were asked to rate the pupils of this latter group on a number of qualities or conditions, including health, nutrition, freedom from physical defect, physical maturity, regularity of attendance, personal appearance, neatness, social status of home, deportment, initiative, originality, responsiveness to suggestion, industry, perseverance, interest in studies, reading

and arithmetical ability, understanding of English language, etc., and also to estimate on a scale of 1 to 5 the child's general intelligence, using the formula previously quoted. Only about a dozen of these have thus far been returned, and the information there given, while in general very illuminating, is not in such form as to permit tabulation.

The number of pupils in each room whose ability the teacher considered to "vary materially" from the score made varied from none or one or two, up to seven or eight, and averaged about four or five per room. In several cases the intelligence-estimate by the teacher proved to vary from the score by not more than one or two deciles. This may mean either that such a degree of variation was considered "material," or that the score fell below or above that of some other individual with whom the teacher was accustomed to compare the child in question, or that the effort at analysis had the effect of modifying the teacher's original impression by the time she came to formulate an estimate of her own. Pupils who were estimated high by the teacher invariably were assigned high standing in industry, perseverance, deportment, and other non-intellectual traits, while most or all of the children estimated low in ability were considered low in the volitional traits. This, of course, does not warrant our assuming that the teacher is misled in her estimates of intelligence; indeed, there is usually a considerable correlation between desirable volitional traits and intelligence; but it does reveal the presence, in these few cases of disagreement, of the possible explanatory causes hereinbefore mentioned.

While the form of the estimates does not permit calculation of coefficients for these groups, the disparity certainly averages no higher, and in many cases much lower, than in the preliminary series, which showed a reasonable conformity of the test scores to other criteria of intelligence.

Analysis of the Scale

What components of general intelligence do the various tests in the scale measure? Or do they all measure the same thing?

In an effort to answer this question, correlations were computed

between the various tests, for six school groups, totalling 226 children.

Right Answers with Opposites	.33
Right Answers with Series Completion	.43
Right Answers with Form	.40
Opposites with Series Completion	.34
Opposites with Form	.30
Series Completion with Form	.46
Probable Error for each coefficient	±.04

Since all the tests correlate positively with each other, it is evident that all test some common ability, or at least abilities which have a common origin. What this common element is can be determined only by an analysis of the tests themselves.

One common component is the ability to associate verbal directions with pictorial forms and with motor acts. To reduce to a minimum the dependence of successful performance on an understanding of English words, the instructions are supplemented with dramatic demonstration of the procedure required. But in the Right Answers test the language factor still remains an essential one. That limited ability to understand English, however, is not an absolute bar to the usefulness of the test is apparent from the scores obtained in School J, in the Italian neighborhood. Rarely is there found a home in this neighborhood using any other language but Italian. Acquaintance with English is practically limited to the school and playground experience. In the beginning class of the first grade there are many children who speak no English, and numerous zero scores are made in this grade. This may be due to language defect, or to inability to meet other conditions of the test, such as sustained attention, etc. The writer's experience in giving the test inclines him to consider the latter a large factor. That the language defect is not the only cause for such low scores seems evident from the fact that the form and direction of the median curve from grade to grade resembles closely the curve for schools where the language difficulty is not present. The J school curve is, of course, very much lower than those of other schools throughout its length, but the difference between the curve for this school

and that of the E system, for example, does not decrease in the third and fourth grades where the language difficulty has been overcome. The lower score in this school cannot be entirely due, therefore, to language difficulties, but is evidently a consequence either of difference in native endowment of this particular stock, or to environmental causes, such as defective nutrition and home care, which are all too common in this community.

Besides the language factor, the four tests have in common certain attention elements, such as ability to hold the "mental set" induced by the instructions until the responses are made. As will be shown later, the type of attention required varies from test to test. Ability to inhibit response pending decision is also involved. Akin to this is the attitude of revision or correction of judgments, the "self-criticism" factor. A certain very limited degree of motor control is involved, although both in devising and scoring the tests, measures have been taken to reduce this to a minimum.

But that the four tests do not measure the same thing throughout is evident from the variation between the coefficients. Let us analyze the tests to see what these differences signify.

The Right Answers test group measures a variety of abilities, although all have certain elements in common, viz.: apprehension of meaning of a concrete situation verbally described, frequent redirection of attention, discriminative reactions, as well as such more specialized abilities as judgment of means to end (No's. 2, 3, 6, 8), spatial judgments (4, 5, 7, 8), a certain amount of information about common objects (2, 3, 6, 9), etc. It is interesting to note the difference in the content of information revealed in such a simple test as this. In school J (Italian), first grade children almost without exception failed to make any response, right or wrong, on No. 6 of this test, an arrow being an object totally outside their experience and vocabulary, although this test occasioned no such difficulty in any other first grade groups. In No. 9 a very frequent response to the instruction to put a circle around the church and a cross on the factory was to put a cross on the steeple of the church, surely a not inexplicable response in a neighborhood where every church bears such a symbol.

The Right Answers test resembles the Series Completion test

(.43) more than it does the Form test (.40) or Opposites (.33). They involve in common a diversity of attack on the successive elements in the test which is not characteristic of either the Opposites or Form tests. The latter tests require a definite preliminary "set," which has to be maintained throughout the minute allotted to the test, while the Right Answers test (and to a lesser degree the Completion test) call for a response whose form, as well as spatial location, varies from one element to the next. That is, the Right Answers and Completion tests require constant motor readaptations. For some, this proves easier than the maintenance of a fixed set; for others, it is probably more confusing.

The Series Completion and Form tests, which show a definite correlation (.46), resemble one another closely in that both test particularly the ability to perceive space relationships (form, size, direction), since more of the elements in the Completion test involve spatial relations than numerical relations. There are, however, distinct differences between them. One difference is in the mental set (already mentioned) with which one attacks them. In the Form test, as in the Opposites, the attention set up by the instructions calls merely for attention to the perceptual-analytic factor,—the form of the blocks to be fitted together,—the response being more or less automatic. Attention is maintained more tensely focussed than in the Completion tests, where it alternates between perceptual-analytic processes and motor response. Hence, the tendency found frequently in younger children to start well but quickly degenerate into a stereotyped response occurs more frequently in the Opposites and Form tests than in the Series completion, and is indicative of failure to hold attention on the instructions given.

But besides a difference in the type of attention demanded in the Form and Completion tests, there is clearly a difference in the mental processes involved. The Form test requires a comparison of forms (and in a few cases, of size), but involves only a selection of one among four which are already present. The Completion test, on the other hand, requires not only comparison of forms, but two other acts not involved in the Form test; first, the abstraction of the significant relationship element which determines the nature of this series; and second, the constructive embodiment of this rela-

tionship in a mark of definite size or form or both. Failures in this test appear to be rarely or never due to lack of motor ability, but either to inability to perceive the relational element (revealed in the omission of certain tests), or else failure to get the *significant* element, revealed in a response which bears some of the characteristics of correctness, but lacks others, as for example, marks of the right shape but wrong size.

The Completion test and Opposites test are the two single tests which, in the preliminary tryout, showed the highest correlation with mental age (.58). But that they do not test exactly the same ability seems apparent, not only from the moderate degree of correlation (.34) between them, but from an analysis of the tests themselves. The difference in type of attention involved has already been described. What has just been said about the mental processes involved in the Completion test is of significance in a comparison with the Opposites test. Both in common involve ability to abstract from each group of pictures or forms some relational element, an ability, certainly, which is one of the largest components of general intelligence as we daily recognize it in people. But beyond this, the two tests call for quite different types of mental reaction. Whereas, in the Completion test, the abstracted element has to be embodied by the child in a definite kind of mark, in the Opposites test a contrast-association response must follow; but because of the difficulty in drawing the contrasting object (and scoring such drawings) the response is made to take the form of selection of one from a group of possibilities, and a simple mark of designation of the one selected. Thus, the motor element in the Opposites test is considerably less than in the Completion test. Analysis of papers shows errors in the Opposites test to fall in one of three general classes: first, those due to inability to get or hold instructions in mind long enough to perform the complicated mental act required, manifesting itself either in a wholly stereotyped response or in a response correct at first but quickly degenerating into stereotyped form; second, apparent inability to make the required abstraction and contrast-association, in which a non-stereotyped but wholly unintelligent series of marks is made (often, of course, it is difficult or impossible to determine which of these two causes is operat-

ing); third, an intelligent attack, but with occasional errors due to making a wrong abstraction, and hence a wrong response. Thus, the child may mark the horizontal arrow instead of the inverted arrow, or the diagonal straight line instead of the curved line, both these errors occurring occasionally, although infrequently, in the papers of children making high scores.

In the opinion of the writer, the value of the series of Right Answers tests is twofold. The correlation coefficient shows that it has considerable significance as a test of intelligence when taken alone, but much more when taken in conjunction with other tests. Thus, its addition to the other three tests of the scale produces a higher correlation with mental age (see Table III) than is obtained by the three tests alone. Besides this, its character makes it peculiarly fitted for an introductory test. It has variety to awaken and sustain interest. Its constant incentive to quick response promotes an attitude favorable to good scores in the other tests. It helps the child get acquainted with the examiner and his ways. It gives him preliminary practice in making responses by marking. It minimizes the penalty for failure to "warm up" to the test, in that the child loses only three points for failure to get himself adjusted to the instructions, where such failure on one of the other parts would mean much greater loss. There are probably fewer failures on later parts than there would be were it not for this preliminary test, although we do not have data to verify this belief.

By what psychological method the individual child goes about making his responses, whether through visual, motor, or vocal imagery, we do not, of course, have data to determine. In the author's mind, there is no doubt that a variety of methods are used. Vocalization doubtless is of much assistance, especially in the Opposites test, where it is distinctly advantageous to have names for the different objects pictured, and where lip movements, or even whispered articulations are constantly noticeable. On the various Right Answers tests, children are frequently observed to make preliminary tentative finger responses, and when satisfied, repeat and record the response with pencil. In the ball-on-table test, children occasionally draw a line showing the path of the ball.

An examination of the papers reveals another significant fact,

namely, that in papers of children of high score, one frequently finds erasures and corrections, indicating criticism and revision of the first hasty judgments. On the other hand, the examiner comes soon to recognize a type of child, of mediocre score, who rushes through the test, finishes ahead of the others, and then waits for time to be called, without taking any pains to improve his efforts. This is nothing more than saying, as Binet said, that one characteristic of general intelligence is the power of self-criticism.

It is interesting to note how the other two characteristics of general intelligence, as Binet conceived it, are revealed by an analysis of this test. The tendency of intelligence to take and maintain a definite direction is revealed in marked fashion in the way in which the abler child holds in mind the instructions for such tests as the Opposites, while the child of less ability frequently forgets them after making one or two correctly. The "capacity to make adaptations for the purpose of attaining a desired end" is, of course, the ability which leads to the discriminative and selective acts and the varied types of response which the various tests demand.

It seems evident that no one of these tests,—and indeed, no single brief test of any sort,—can measure all phases of that composite general ability which we call general intelligence. The adaptations which the intelligent person makes in the effort to attain his ends are highly varied in character to meet the varying materials with which he works and the varying ends which he seeks; and they require very diverse types of mental activity, even in the earlier years of life. The highest test scores are not made by the same combination of actions, which is to say, equally intelligent people do not attack their problems in exactly the same way, nor show abilities equal in every respect. Hence, a variety of tests must be used in order to touch the several phases of general intelligence.

How many tests are necessary for this purpose is a matter of dispute. It seems to be the belief of Thorndike, Otis, and others, embodied in such scales as the Army Alpha Tests, the Columbia College Entrance Tests, etc., that there must be many tests (since there are so many special abilities that go to make up general ability). This may, in some cases, make for slightly higher co-

efficients of correlation than scales comprising fewer tests. But it is open to question whether, after all, such elaborate forms of scale, with all their refinements of statistical method, are in practice preferable to briefer and more easily administered tests. The value of the long, elaborate scale seems greater in the upper years where test makers have generally found their greatest difficulties. But in the earlier years, where the variety of mental operations by which the child makes his relatively simple adaptations is so much smaller and less complex, there is reason to question the superiority of such a type of scale. Particularly is this true with group tests, where the physical limitations of childhood make a brief test practically indispensable.

Of significant bearing on this problem are the coefficients of correlation cited in Table III, which show that a group of six tests is not superior as a measure of mental ability to any combination of four tests chosen from among them. We are probably not warranted in concluding that six are inferior to four, save as they would lengthen the test period and make for fatigue and loss of interest. The six tests on which the correlations are based were, be it remembered, given in two periods of three tests each, about twenty-minute periods, so fatigue and waning interest did not operate to reduce the scores made. Nevertheless, the validity for the six combined is slightly lower than that of four, if the coefficients mean what they appear to mean.

This is by no means to say that a single scale of this sort provides as valid and reliable a measure of general intelligence as can be obtained. The writer's belief is distinctly to the contrary. He hopes that much more adequate scales than this may be devised, if they have not already been devised. As was suggested earlier, one of the urgent immediate needs is a thoroughgoing comparative study of different types of test material and scales for group testing of primary children, as well as for children of higher grades.

But beyond this, no single test will permit every child to make as high a score as his ability warrants, because not every child is in the most favorable physical and mental condition at the same time. Moreover, as has been previously suggested, the group method of testing, with all its conveniences and economy, prob-

ably can never be so reliable a measure of a child's ability as an individual test, where the examiner can take whatever time and measures are necessary to get a maximum of response from the child. A distraction during the giving of a group test instructions may cause a child to miss an essential point and lower his score. It is, therefore, an undesirable practice to make the child's pedagogical or social future dependent on the results of a single twenty- or thirty- or even sixty-minute test. Our analysis of the psychological factors which make for success or failure in life is still too theoretical to warrant staking so much on a single score figure. The greatest danger to the mental testing movement is from the over-enthusiasm of its undiscriminating advocates, which sooner or later must lead to a disastrous reaction against the whole educational measurement movement, if not tempered by conservative judgment.

The test score, nevertheless, has its distinct and valuable function, and that is, to aid the teacher or parent or other responsible adult to a correct diagnosis of the child's actual ability. After all, no disposition of a child's case, either by irregular promotion, nonpromotion with class, assignment to special class, school, or institution, should be made save as a result of an extended, careful, intelligent, and sympathetic diagnosis of the child's whole personality, character, and abilities, both special and general. While no test score can ever serve as a reliable substitute for this, a good test score can be of material aid in making such diagnosis, by confirming the teacher's judgment, or by contradicting it one way or the other and leading to a more careful and truer analysis. For such purposes, evidently, two or more test scores, obtained at different times and under varying conditions would be better than one, however good that one might be. It would seem, then, especially after the development of group scales making possible larger economies, that a testing program in any school should call for several intelligence tests, especially during the early years of the child's school career, supplemented by such physical, educational, volitional, and other tests as are useful.

APPENDIX I

Additional Right Answers Tests Used

The illustrations (Plates I to VI) show the tests used in this investigation. Plates I to IV show the final form of the scale. The preliminary group form included the four parts of the final scale (I. Right Answers, II. Opposites, III. Series Completion, IV. Form test), and in addition, V. Associated Objects test (Plate V) and VI. Analogies test (Plate VI).

The Right Answers test used in the preliminary group form included also eight additional tests, listed in Table 2, which were eliminated and not counted in computing the correlations in which the Right Answers test appears. These eight tests are not illustrated here, not being considered worth further investigation in their present form, but they may be briefly described and their instructions summarized, as follows:

No. 11. "Tools." Pictures of square, drawknife, saw, compass, hammer. "Which tool does the carpenter use to drive nails?"

No. 12. "Toys." Pictures of trumpet, drum, bicycle, dog, toolbox. "Which of these Christmas presents can Robert ride?"

No. 13. "Animals to Eat." Pictures of dog, rabbit, frog, horse, owl. "Which animal is best for the hungry man, lost in the woods, to catch and cook for his dinner?"

No. 14. "Rectangle." Rectangle 1″ x 2″. "Draw a line which will divide this oblong into two squares." Eliminated early because too easy, unless scoring standard were made too rigid to be practicable or desirable.

No. 15. "Heavy and Light Objects." Pictures of hand-axe, feather, book, basket, pencil. "Draw a line around the heaviest object, and a line under the lightest object." Eliminated early, because instructions were found to suggest response too readily; also because some uncertainty whether basket pictured would not be as heavy as some handaxes.

No, 16. "Animals." Same picture as demonstration test, No. 1. "Which animal frightened the man?"

No. 17. "Tools to Cut." Pictures of shears, hatchet, saw, butcher-knife, pocket-knife. "Which tool is best to use in cutting a board to mend the fence?"

APPENDIX II

Instructions for Giving Primary Group Test

(Examiner should familiarize himself thoroughly with instructions before undertaking to give the test.)

PRELIMINARY

See that all pupils have pencils. Desks should be cleared. Ascertain in advance the method of collecting papers customarily used in the room, so that after the test the papers can be taken up without delay. Say:

Now we are going to play a sort of a little game together, with pencils and pictures. The pictures are in these folders I am going to pass you. I want you to leave them lying on your desks, and not touch them, or turn the pages, or do anything to them, until I have told you just what to do.

It will be necessary throughout to guard carefully against children's marking when they should not, or looking ahead while instructions are being given. Do not hesitate at any moment to remind them, "Don't mark till I tell you," or, "Everybody lay your pencils down," or, "Look at me," etc. For the same reason, have papers taken up as quickly as possible after the test is finished.

Distribute folders, laying them face up on desks. Warn pupils again not to open them or turn pages. When all are distributed, hold one up and, pointing to name line, ask pupils to write their names on the line; then, pointing to age-line, says, "Put here the figure that tells how old you are." If the children cannot write their names, it should be done by the teacher before folders are distributed.

The other data called for on page 1 may be supplied by pupils at this time, if time and ability to write permit. In the first grade, and often in higher grades, it is more economical of time to omit this and supply data at the examiner's convenience from records furnished by the teacher or noted on one of the folders by himself. In this case he should enter on each folder used a "key" number or symbol, in the space at right marked "Lot No.," to designate all folders used in a given room at one time. If more than one examiner is giving tests, the key number may be prefixed with the examiner's initial; thus, "J-1" may mean the first test given by Miss Johnson; "B-12," the twelfth test given by Miss Brown. Because of children's difficulties in spelling names of months, as well as because fractional parts of a year have not been used in norms, birthday has not been asked for. If this datum is desired, it can be secured from school records and entered in right-hand blank marked "C.A."—chronological age—in years and months. Blanks are also provided for recording date of giving the test, child's mental age, intelligence quotient, and teacher's estimate of his intelligence, if these are available. In some cases it has been found necessary to check up pupils' statements of their ages with the school records, particularly in neighborhoods where many children leave school at the earliest

legal age, and a tendency to overstate the child's age is prevalent in the community.

It is extremely important that the time allowance for each test be rigidly observed. Use stop-watch if possible; if not, use second-hand on watch. Do not allow one second too much or too little time. Failure to keep time allowance with precision destroys the possibility of comparison of scores with other groups, and may invalidate comparison within the group itself, by allowing slow children to continue and catch up after quick children have finished, thus concealing differences which really exist. Don't expect pupils to get through in time allowed, or even to get far along. The time has purposely been made so short that very few third-grade children can finish parts II, III, or IV in the time allowed.

On Part I (Right Answers) the examiner should have some preliminary practice on counting seconds so that he can time these tests accurately and easily without looking at his watch. One suggested device is to repeat silently, "one-chimpanzee, two-chimpanzee, three-chimpanzee, four-chimpanzee," etc., or use some other polysyllabic word which cannot be enunciated too rapidly. Practice with a watch, before undertaking to give the test, for several fifteen or thirty-second periods, to get the counting rate uniform and natural. Time should be counted from the last word of instruction for each of the tests in Part I.

After blanks have been filled by pupils, say:

Now lay your pencils down, look at me, and listen carefully while I tell you what we are going to do. In these folders are some pictures and drawings; we are all going to do certain things with them; I will tell you just what to do, and I want to see if you can do *exactly* what I tell you, and *how quickly* you can do it. Now nobody must look to see what anybody else is doing, because that wouldn't be fair. We want to see what you can do all by yourself. Listen very carefully to all I say, so you will be sure to hear the first time. Don't turn any pages until I tell you to. And don't ask me any questions; if there are any things too hard, or that you don't understand, skip them and do the rest. But just do the *very best* you can, and do the things as *quickly* as you can without making any mistakes. That is what the game is.

Now we will all look at this first page.

I. *Right Answers Test*

I am going to tell you some little stories and ask you some questions, and I want you to answer them by marking the right picture in the way I shall tell you. First I will tell you one and mark it, so you can see how.

Look at number 1, up in this corner (point).

1. As I walked through the woods one day, I saw these five animals. One was flying from tree to tree. Draw a circle around the picture of the one that was flying from tree to tree.

Which should it be? Why, the bird, of course, because that is the only one that can fly. Take your pencils and mark a circle around the bird, like this (mark). It doesn't have to be a good circle, just a ring around it, so I can tell that you know which one is right. Now don't mark any more till I tell you. As I tell you the rest, you will have to work as quickly as you can without making a mistake. Don't answer aloud, but just answer by marking in the right place.

Look at number 2, the hats in the next picture (point).

2. When the weather turned cold, a man went to the clothes closet to find what to wear on his head. Here are the hats and caps he found. Which is best to wear on a very cold day? Draw a circle around it. (Five seconds)

Look at number 3, this next picture (point).

3. Helen's mother left her at home to take care of the baby. The baby cried, so Helen went to find something for the baby to play with. Which of these things may Helen let the baby play with? Draw a circle around the *two* things the baby may have. (Eight seconds)

Look at number 4, at the beginning of the next row; this lot of circles (point).

4. Draw a line from the largest circle to the smallest circle. (Eight seconds.)

Look at number 5, these dots in the next picture (point).

5. Henry was on his way home from school, when he dropped seven pennies on the sidewalk. Here they are, lying just as they fell; each dot is one penny. Before he picked them up, he said, "Look! here are three of the pennies in a straight row!" Draw a straight line touching the three pennies that are in a row. (Ten seconds)

Look at number 6, the next picture, right beside it (point).

6. Finish the arrow. (Five seconds)

Look at number 7, down in this lower corner, the table (point).

7. Here is a table, with a ball on the edge of it, and some blocks on the floor below. If the ball should roll off the table, which block would it strike? Mark a cross on the block it will strike. (Five seconds)

Look at number 8, next picture, middle of the bottom row (point).

8. Here is Francis, just starting for school. You see four paths leading from Francis to the school-house. It is almost school time, and Francis will have to take the shortest path and hurry to keep from being late. Which path should

he take to be sure to get to school on time? Mark a line along the path he should take. (Eight seconds)

Look at number 9, these buildings down in the corner (point).

9. Mr. Brown is a minister, or preacher (or "priest," in some communities); Mr. Jones runs a machine which makes furniture. Here are pictures of the buildings in which they work, and some other buildings besides. Draw a circle around the picture of the building where the minister (priest) works, and put a cross on the one where the furniture maker works; a *circle* around the one where the minister works, a *cross* on the one where the furniture maker works. (Eight seconds)

Now everybody stop, put your pencils down, and *watch me.* (Wait till you see all have laid pencils down.) Now, turn the page, and fold it backward, like this (demonstrate). Then lay it down so *this* page of pictures (Opposites test) is on top.

(It is well to pass rapidly along aisles as children are turning the pages and see that each child has the right page on top; some confuse it with the Completion test.)

II. *Opposites Test*

Here are some queer pictures, several rows of them, with one picture in the square at the beginning of each row. You see there are two sets of rows on this page, these (point) and these (point), so each row goes only half way across the page. Now in each row there is one picture which is *more different* from the one in the square than any of the others. That is the same as saying it is the *opposite* of the one in the square. I wonder if we can find it in each row? Look here, and I'll show you what I mean.

See, in this first row (point) we have a *long line* in the square, then three shorter lines in the row (point to each). This middle one (point) is the *shortest,* so it is the *most different* from the one in the square, which is the longest. So I take my pencil and draw a circle around it (mark), because it is *most different,* or *opposite,* from the one in the square. Now everybody take your pencils and draw a circle around this shortest line, the middle one. Now lay your pencils down. (See that all lay pencils down.)

Now look at the second row. See, in this square (point) we have a *black spot;* then after it come three other spots. Which of these other three spots is most different from the black spot? Why, this *white* one (point to last), of course, because *white* is the *opposite*

of *black*. So I mark a circle around it (mark). Everybody take pencils and draw a circle around the white spot. Now lay your pencils down and look at me.

Now listen carefully. After I tell you to begin, I want you to look carefully at the picture in the square in the next row, then at the three pictures after it, and see which one is *most different,* or *opposite,* from the one in the square, and draw a circle around it, just as we did with these others. Don't mark until I tell you; leave your pencils on the desk. Then do the same with the rest. Do all on this page, both columns, these (point to first column) and these (point to second column). If you find any you can't do, skip them. Work as fast as you can without making any mistakes.

Ready, take pencils, BEGIN.

(One minute)

STOP. Lay your pencils down, everybody. (See that all stop.) Now watch me; turn your folder right over (demonstrate) like this, so this (hold up page 3, "Completion," toward pupils) will be on top. (Some children will start to turn page instead of folder, so be on guard against this.)

III. *Series Completion Test*

Here are several rows of drawings of different kinds. In each row there is a blank space, sometimes two blank spaces, where a drawing has been left out which is needed to finish the row. I wonder if we can figure out what is left out in each row?

Look carefully at this first row (point). Here are a lot of straight lines, one after the other, all alike, *except* that they get *shorter* as we go across the page. This (point) is the longest; this (point to each in turn) is shorter, this still shorter, shorter, shorter, shorter,—and then the blank space. Now what should we put in the blank space to finish the row nicely? Why, a little, *short* line, shorter than any of the others, like this (mark). Everybody take pencils and mark one on your paper in this blank space. Now the row is finished. Lay your pencils down.

Now look at this second row. Here (point, successively) we have an O, and X, O, X, O, *blank space,* and O. What should go in the blank space? Why, an X, of course, because we skipped the

X that goes between the O's. So we put an X here (mark), and the line is finished. Take pencils and mark an X on your folder. Now lay down your pencils and look at me.

After I tell you to begin, I want you to look carefully at the drawings in each row, make up your mind what is left out from the blank space, or blank spaces, then take your pencil and draw in whatever has been left out, so the row will be finished. Don't mark until I tell you. Do all on this page, both columns, these (point), and these (point). If there are any you can't do, skip them. Work as fast as you can without making any mistakes.

Ready, take pencils, BEGIN.

(One and one-half minutes)

STOP. Everybody lay your pencils down. Watch me. Now, turn the page over and fold it back like this (demonstrate), so this page (hold up fourth page toward pupils) will be on top.

IV. *Form Test*

Here are several rows of drawings we will call *blocks*. You see, in each row, at the beginning of the row (point), a queerly shaped block, followed by four other blocks.

Now, in each row, we want to make the block at the beginning of the row into a solid square block, by fitting up against it, or fitting in with it, one of the *other* blocks in that row, so that together they will make a solid *square* block like this one (point) at the top of the page. I wonder if we can find the right one to use in each row?

Look carefully at this folder I have, and I will show you what I mean. See, in this first row (point), the block at the beginning of the row is three-cornered. Which one of the blocks in the row is shaped so it could be fitted up against this first block to make a square? Why, this last one; none of the others would do; they are not the right shape. (Note: it is advisable to sketch outlines of blocks in this and the other demonstration on the blackboard, as you talk, close to each other, thus:  .) So I draw a circle around this last block (mark), because it is the right one to use. Everybody take pencils and draw a circle around this block. Now lay your pencils down.

Now look at the next row. See this curved block (point), shaped like a letter "D", at the beginning of the row. Now which of these other blocks, one, two three, four (pointing), would fit in with this to make a square? Why, this second one would; see, it is curved *in* (draw on blackboard and point to curves as referred to), so it would fit right against this side which is curved *out* (point), so together they would make a perfect square. So I draw a circle around this second block (mark), the one with the two sharp points on it, because it is the right one to use in this row. Everybody take pencils and draw a circle around it. Now lay your pencils down and look this way.

After I tell you to begin, look carefully at the block at the beginning of each row, so you'll remember its shape; then see which one of the other blocks in that row will fit in with it to make a solid square block; then draw a circle around the right one, as we did with these first two. Don't mark until I tell you. Do all on the page, both columns, these (point), and these (point). If there are any too hard, skip them; work as fast as you can without making mistakes.

Ready, take pencils, BEGIN.

(One minute)

STOP. Everybody lay your pencils down. Now turn your folders over, so your name will be on top, and we will take them up. (Collect folders promptly.)

INSTRUCTIONS for *Associated Objects Test,* Preliminary Series.

Here are some pictures in rows, with one picture in a square at the beginning of each row. Now in each row there are pictures of either one or two things that belong with, or go with, the thing in the square. In some rows there is one, in some rows two; so you must look carefully at each picture to make sure you don't miss anything.

See, in this first row, we have in this square (point), a table-knife. Then in the row we have (pointing to each) a watch, a radiator, a fork, a spoon, and a flag. Now which of these belong with the knife? Why, the fork and spoon, of course; we always find knife, fork, and spoon going together. So we draw a circle

around the fork and one around the spoon (mark). Everybody take pencil and draw a circle around the fork and one around the spoon. Now put your pencils down.

Now look at the second row. In the square we have (point) a soldier. Now which of these things in the row belong with the soldier? Why, the gun, of course. So we put a circle around the gun, like this (mark). Everybody do that. Now lay your pencils down, and look at me.

After I tell you to begin, I want you to look carefully in each row for the one thing or the two things that go with the thing in the square at the beginning of that row, and then mark a circle around those things. Don't mark until I tell you. Look carefully at all the pictures in each row to make sure you don't miss anything. Work as quickly as you can without making mistakes. If there are any you don't understand, skip them.

Ready, take pencils, BEGIN.

(One and one-half minutes)

STOP.

INSTRUCTIONS for *Analogies Test,* Preliminary Series.

Here are several rows of drawings of different shapes. In each row there are three drawings and then a blank space (point, 1, 2, 3, 4). Now what we want to do is this: draw in each blank space something that compares with the third thing in the same way that the second thing compares with the first thing. Look here; something that compares with this (point) in the same way that this (point) compares with this (point). Now listen closely, and watch me, and I will show you what I mean.

Here in this first row we have, first, a large circle (draw figures on blackboard as you talk); second, a small circle; third, a large square; fourth, a blank space. Now how is this first thing (point) *like* this second thing (point)? Why, they are both the same shape, aren't they? Now how are they *different* from each other? Why, they are different in size, aren't they,—one large, the other small? That is what I mean by "compares." Now what is it that compares with this (point), the large square, in the same way that the small circle compares with the large circle? Why, a small

square, of course, same shape, smaller size. So I draw in the blank space a small square, because it compares with the large square in the same way the small circle compares with the large circle. Everybody draw a small square.

Now look at this next row (point). Here we have a U-shaped drawing (draw on blackboard); second, the same thing, but *up-side-down;* third, this sort of square with the top line off; fourth, blank space. Now what can we put in the blank space that compares with this third thing (point) in the same way this second thing (point) compares with the first? Why, this thing (point) turned up-side-down, just as the second U is turned up-side down. Everybody draw one like this (mark). Do you see what I mean?

Now, after I say "Begin," look carefully at the drawings in the next row; then decide what compares with the third thing in the same way the second compares with the first, and draw it in the blank space. Then do the same with all the other rows. If there are any you can't do, skip them.

Ready, take pencils, BEGIN.

(Two minutes)

STOP.

APPENDIX III

Instructions for Scoring Primary Group Test

With aid of scoring key held in left hand against the successive pages of the scale, check responses on the test blank, in right hand margin of Part I, and center and right hand margins (*i.e.,* those at right of each column) in other parts. Check "R" (or √) for right responses, "W" (or X) for wrong responses. Number 1 in Part 1, and the first two in each of the other parts, being used for demonstration purposes, are not scored. The various parts are weighted as follows:

Part 1 (Right Answers). The number of rights is multiplied by three, and the product entered at top of page 1, opposite "1. Rights x 3."

Part 2 (Opposites). Both rights and wrongs should be scored, half the number of wrongs subtracted from the number of rights,

the difference multiplied by two, and entered at top of page 1, opposite "2. (R — ½ W) x 2." This partial deduction of errors is made to offset the one-in-three chance that the child will mark any row correctly by mere guesswork. Thus, if he marked twelve tests by pure chance, or in some stereotyped form of response (*e.g.*, marking first in each row), he would probably get four right and eight wrong, which, corrected as above described, would give him a zero score; whereas a smaller proportion of errors would give him a plus score, although not so large as if he made no errors. A minus score, due to making more than twice as many wrong responses as right, is entered as a zero score. (An alternative method, giving the same result, but avoiding subtraction of fractions, is to multiply the number of rights by two and subtract from it the number of wrongs, entering difference on page 1.)

Part 3 (Series Completion). The number of rights is multiplied by two and entered at top of page 1. No deduction is made for errors, since chance of a successful response by merely guessing is very slight.

Part 4 (Form). The number of rights is multiplied by two and entered at top of page 1. No deduction is made for errors, since chance of successful response by guessing is small (one in four), and has been found to be negligible in practice.

A perfect score is 100, and would be obtained as follows:

Part 1,	8 elements	(omitting number 1)				weighted by	3,	total	24
" 2,	14 "	("		first two)		" "	2,	"	28
" 3,	12 "	("	"	")		" "	2,	"	24
" 4,	12 "	("	"	")		" "	2,	"	24
									100

There are no fractional scores.

DETAILED SCORING INSTRUCTIONS

(Supplementing Key)

Part 1

In all cases where two objects are marked and only one should be (whether right one is one of those marked or not), score W.

In all cases where only one object is marked and two should be (whether one marked is right or not), score W.

No. 4. Any line *connecting* the two circles is R. Merely indicating the two circles is W.

No. 5. Line must *connect* the three correct dots, but need not be straight.

No. 6. Shaft of arrow must run back to notch at left end of feathering, but not beyond, and must run forward to head; need not be straight.

No. 7. Any mark that clearly designates the correct block is R.

No. 8. Any mark that clearly designates correct path is R.

No. 9. Marks designating church and factory must be such as can be liberally interpreted as "circle" and "cross" respectively. Interchanging, omission, or marking any in addition to the two directed, is W.

Parts 2 and 4

Interpret "draw a circle" liberally; any unambiguous mark of designation is acceptable.

If drawing in square (part 2) or at beginning of row (part 4) *alone* is marked, score W. If *both* it and correct response are marked, score R. If any other two (or more) are marked, score W.

Part 3

3rd row. (alternate vertical and horizontal lines). Line need be only approximately horizontal and straight.

4th row. Circle must be larger in *one* dimension than adjacent printed circle, and not less than two-thirds as wide in other dimension.

5th row. Both blanks must be correctly filled.

6th row. Square must be smaller in *both* dimensions than adjacent squares. Consider number of sides, rather than accuracy of shape.

8th row (top of second column). Three dots in any arrangement, or figure "3", in circle.

9th row. Two dots, any arrangement; no circle.

10th row. Circle smaller in at least one dimension than adjacent circles.

11th row. Both blanks must be filled. Number, rather than exact form of marks, counts.

12th row. Line must be vertical, or more nearly vertical than adjacent line, and should be at least half as long as other lines in row.

14th row. Rectangle must be *both* longer and narrower (in at least half its length) than adjacent rectangle.

KEY to *Associated Objects Test,* Preliminary Series

No. 1. Demonstration
2. Demonstration
3. Nail
4. B, C
5. Spade, Hoe
6. Telephone Pole
7. Envelope, Letter
8. Oar
9. Bow
10. Leaf, Apple
11. Mouse-trap
12. Door, Padlock
13. Sign-post, Car
14. Photograph
15. Arrow, Target
16. Corn-stalk

Interpret "draw a circle" liberally.

If only one drawing is marked when two should be, score W.

If more than one drawing (unless one be the drawing in square) is marked when only one should be, score W.

If drawing in square *alone* is marked, score W. If both it and the correct response are marked, score R.

If one right and one wrong drawing are marked, score W.

KEY to *Analogies Test,* Preliminary Series.

No. 1. Demonstration; not scored
2. Demonstration; not scored
3. Large circle
4. Short horizontal line
5. Short vertical line
6. Triangle inverted
7. Circle enclosed in circle
8. x x x
9. Short diagonal line, same direction as drawing 2

10. 5 x's, same arrangement as drawing 3
11. Square or diamond
12. Like drawing 3, but horizontal line to right of vertical
13. Circle, approximately diameter of square
14. Three squares, same arrangement as drawing 3
15. Semicircle (empty)
16. Square, bottom line omitted or indented to center of square.